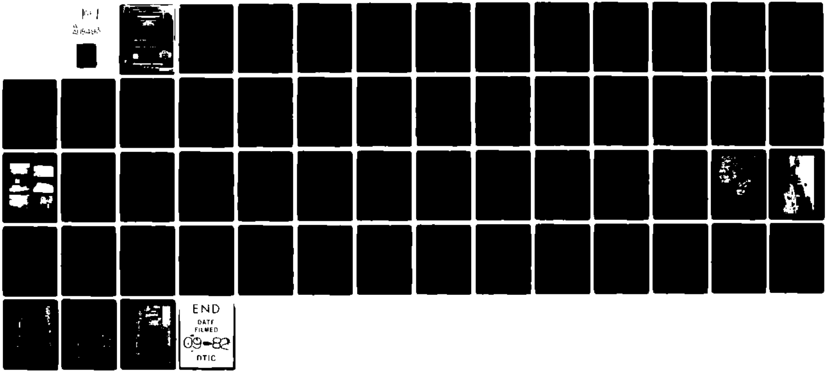
AD-A118 493   CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER --ETC F/G 9/2
WORKING PAPERS IN ACQUISITION OF KNOWLEDGE FOR IMAGE UNDERSTAND--ETC(U)
DEC 76   O AKIN, R REDDY, R OHLANDER, M SCHULTZ   F44620-73-C-0074

UNCLASSIFIED   NL

1 of 1
AD A118493



END
DATE
FILMED
09-82
DTIC

DEPARTMENT

of

COMPUTER SCIENCE

WORKING PAPERS IN ACQUISITION OF KNOWLEDGE FOR IMAGE UNDERSTANDING RESEARCH

Omer Akin*, Raj Reddy, Ronald Ohlander
and Marty Schultz
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

December 15, 1976

* Also with the Department of Architecture. This research has been sponsored in part by the Defense Advanced Projects Agency under contract no. F44620-73-C-0074 and monitored by the Air Force Office of Scientific Research.

CONTENTS

ABSTRACT

Use of knowledge has facilitated complex problem solving in many areas of research. However, in the Image Understanding area, we do not have any systematic treatment and codification of knowledge that is useful in image perception. Further, we do not even have adequate tools for acquiring the necessary knowledge base. In this report we present an experimental paradigm for knowledge acquisition, discuss an analysis technique, and illustrate the different types of knowledge that seem to be useful in image understanding research.

In the first paper, three major aspects of knowledge are presented: primitive Feature Extraction Operators, Rewriting Rules, and Flow of Control. A limited number of Feature Extraction Operators were repeatedly used by the subjects to specify *location*, *size*, *shape*, *quantity*, *color*, *texture*, and *patterns*, of various *components* found in scenes. Six types of Rewriting Rules were identified; *assertions*, *negative assertions*, *context-free*, *conditional*, *generative*, and *analytical* inferences. Flow of Control exhibited characteristics of an hypothesize and test paradigm capable of using *imprecise*, *conflicting* hypotheses in *cooperation* with others in a *multi-dimensional* problem space.

The second paper discusses the picture-puzzle paradigm and the various ways in which it can be used as a tool for acquisition of knowledge. The third paper deals with a computer program that assists the transcription of typical protocols obtained from the picture puzzle tasks. Finally, the last paper of the report discusses the pros and cons of using eye-fixation data to acquire knowledge used in some of the tasks of the picture-puzzle paradigm.

The total effort represents an account of the initial results of a new experimental paradigm. We hope that this will provide a sound basis for understanding the issues of knowledge used in visual perception and aid in the modelling of "seeing" systems.

Knowledge Acquisition for Image Understanding Research

Omer Akin* and Raj Reddy
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

I. INTRODUCTION

Most researchers believe that knowledge based systems will permit significant advances in the analysis, description, and interpretation of images. The fact that knowledge can be used to constrain search has been demonstrated in several other areas such as chemistry and speech understanding. In the past, researchers have used terms such as "linguistic approach" to indicate the desirability of using known structural relationships. More recently the term "semantics" has been used to represent the knowledge based aspects of image processing research.

In general, however, knowledge comes in all sizes and shapes and the use of all such such knowledge can be helpful in the analysis and description of images. We call this process Image Understanding as being distinct from scene analysis and pattern recognition. (Reddy and Newell, 1975) The main contributions of this paper are to present an experimental paradigm for knowledge acquisition, discuss an analysis technique, and illustrate different types of knowledge that seem to be useful in image understanding research.

Several earlier systems have attempted to use problem specific knowledge in the analysis of scenes. Guzman (1968) has used knowledge based on spatial relationships. Although his system is capable of dealing with complex scenes, its performance is limited to planar surfaced objects. The kinds of spatial relations used by Guzman also appear as a part of our results and will later be discussed under knowledge about useful primitive features.

Kelly (1970) has used a specialized knowledge representation to recognize pictures of people. He used specific, salient features of different parts of the human body to detect them. In this sense, he developed a predefined knowledge base for his specific area of application, limited by the special requirements of the recognition task.

Waltz (1972) attempts to use knowledge represented as constraints. In addition to the spatial relations used in identifying regions, he uses prior knowledge about shadows and occlusions to process complex scenes.

Most of these studies tend to deal with knowledge in a task-specific, specialized manner. This paper presents a general paradigm of research and permits generalizations across domain dependent knowledge in a systematic way.

* Also with the Department of Architecture.

Turning our attention to the other related area of research, cognitive psychology, we find that .the research contributions there are not very helpful either. Research on vision provides an abundance of psychometric information about human vision (Julesz, 1971; Hochberg, 1964, 1968). Recent studies in cognitive psychology have accumulated considerable substance about the information processing aspects of perception (Farley, 1974; Moran, 1973; Baylor, 1971). However, the content and role of knowledge in visual perception has been almost completely neglected. There are three major reasons for this which are directly related to the tradition of research in experimental psychology.

For one thing, in the standard psychology experiment the knowledge available to subjects is generally an issue to be controlled rather than investigated. This is largely why almost all major studies in vision, with the exception of a few (Buswell, 1935; Shepard, 1976) deal with abstract stimuli with measurable information content.

Secondly, the usual measures used to calibrate the independent variables are not suitable to measure knowledge. These measures are either based on reaction times or eye-fixation data (Buswell, 1935; Shepard, 1976; Loftus, 1974). The obvious logical explanation for this is that the ultimate goal of all of these efforts is to develop models for perception where the calibration of processing parameters is of utmost importance. Recently protocol analysis, though potentially useful for investigating issues of knowledge, has been used towards the same ends.

Finally, no adequate tools of analysis are available to interpret and codify the data obtained such that issues of knowledge can be dealt with directly. This is partly due to a lack of interest in codifying knowledge specifically, in the area of vision.

On the other hand, the accumulation of the findings of previous research in psychology has contributed to our present ability to deal with the knowledge acquisition issue. From the studies on the processes of visual perception we derive the existence of special mechanisms for inference making and selective processing. In studies with eye-fixations and studies with simple, abstract stimuli there is complementary evidence to the availability of special operators for extraction of visual features. We hope that these studies and the tool proposed here for exploring the knowledge acquisition issue will be complementary to one another.

The research tool proposed here is similar to Newell's (1968) protocol analysis method and Woods and Makhoul's (1974) simulations. Our "picture-puzzle" paradigm consists of providing a man-machine system which simulates a semi-visual-and-semi-verbal channel that can transmit information to subjects about a given visual scene, when requested. The protocols we analyzed were obtained from the picture-puzzle task. The analysis consists of scanning the protocols for the occurrence of different kinds of knowledge sources used by the subjects.

This basically is very similar to protocol analysis in the usual sense. However, in this case a detailed description of the problem states and the problem behavior graph is not necessary for the analysis. We merely inspect the protocols for repeated patterns of utterances and behavior without a real need to formalize the problem space. So, this study differs from Woods and Makhoul's simulation in that it attempts to explore what knowledge sources there are rather than investigating how predefined sources of knowledge cooperate in a given task environment.

Another study by Furschein and Fischler (1972) comes closer to the picture-puzzle paradigm than any other. They have collected protocols on subjects verbally describing scenes, after examining them visually. The control variable they use is the purpose of the description, which covers things like description of scenes for a general purpose data-base or for a city-planning data-base. The analysis focuses on the content and syntax of the scene descriptions provided by the subjects while ours is concerned with the knowledge and mechanisms useful in generating these transcriptions in the first place.

II. METHOD

The picture-puzzle paradigm used was human simulation of image understanding under conditions similar to machine perception. Each subject was asked to find out the contents of a color photograph of a scene without visually seeing it. The subjects were allowed to ask questions about the scene and the experimenter answered these questions using the actual photograph as reference. The questions were limited to the lower level attributes of the scene. By lower level we mean information that does not specify object concepts but properties of segments and regions such as location, color, shape etc.

The picture-puzzle paradigm has three advantages for the purposes of this study: a) the phenomenon of visual perception has been removed from the status of a spontaneous (uncontrolled) human behavior and placed in the status of problem solving, just as in the case of computer vision, b) the visual perception process has been slowed down by several orders of magnitude, c) the interaction of the image and the subject is channeled via the experimenter so that it can be recorded in the form of a protocol.

A. EXPERIMENTAL SET-UP

Two video terminals used in the experiments were connected at the onset of the experimental session by means of software (TALKK program) to enable typed-in communication between their users. The teletypes were located such that no visual communication between their users was possible (Figure 7). The subject and the experimenter were able to communicate only verbally thru the TALKK program.

TALKK was designed to record all statements made by both the subject and the experimenter throughout an experiment. It also enabled them to input conjectures and notes about the task at any time during the experiment. It further enabled subjects to correlate their personal notes and drawings that they were allowed to make on separate sheets of paper, with the protocol. A more detailed description of the system is given in Ohlander, Reddy and Akin (1976).

B. SUBJECTS

The main objective of this study is to observe the knowledge used by humans in visual understanding. In order to obtain a "good" sample of instance of such knowledge we selected subjects that were superior to the average individual in some relevant respect. Four of the six subjects used were knowledgeable in information processing and/or visual perception. The remaining two were architects by profession and had considerable practice in solving complex visual problems. All subjects were college graduates with graduate education ranging through the Ph.D. level. Hence, the subjects were a priori assumed to have considerable expertise in visual information processing.

C. STIMULI AND THE TASK

The stimuli used were produced for and used in automated image understanding research by Ohlander (1975). All the scenes were constructed or selected as usual natural images of familiar objects. Figures 1 through 6 contain the images used.

The subjects were simply instructed to *understand* the contents of a stimulus so that they would be able to describe all major objects in it immediately after the experiment. The experiments were terminated when the subjects thought they *understood* all major objects in the scene or at the end of 2 and 1/2 hours, which ever came first. Only one subject continued with the experiment after the 2-1/2 hours.

The subjects were required to perform the picture-puzzle task with any one of the six different stimulus scenes. At no time during the experiments were the subjects allowed to see these photographs . However they were required to ask questions about them to the experimenter, who at all times had the photographs available for his visual examination.

All verbalizations from the protocols were automatically recorded by the software used. Five types of entries constituted a protocol: questions, conjectures and personal notes of subject; and answers and notes by the experimenter. At the end of each session subjects were asked to recognize the stimulus picture among 19 other pictures some of which resembled the stimulus in terms of content and all of which resembled it in terms of color and print quality.

III. ANALYSIS AND DISCUSSION OF PROTOCOLS

A priori we have partitioned the knowledge used in image understanding into three broad categories. First there is a need for operators to define the various physical attributes (or features) in a stimulus of visual kind. Colors, shapes, locations, orientations and textures of objects are properties that can be easily abstracted and they are integral parts of the knowledge we use to understand scenes. We call this set of visual concepts, Feature Extraction Operators.

The second category of knowledge relevant to image understanding has to do with how we translate the visual information captured by the Feature Extraction Operators into meaningful physical objects or images. Suppose we look down and see a green textured surface under our feet. How do we know what that surface is? Given all the additional knowledge about our physical context (fresh air of outdoors, a supporting surface under our feet, etc.) and the visual impulses we get from the surface (color, texture, etc.) we infer that we are standing on grass. Hence using the temporal (outdoor, support, etc.) as well as the general (color, texture of grass) knowledge we conclude that the surface must be a grassy surface. The knowledge that enables us to translate individual features (context, color, texture) into object concepts (i.e., grass) is called Rewriting Rules.

Finally we use the Feature Extraction Operators and the Rewriting Rules deliberately, or in a nonrandom fashion, to generate the desirable conclusions. For instance, in the above example the desirable result is to identify that the surface under our feet is grass. First we extract certain features from the environment-- such as green, textured, horizontal solid surface, etc. Next, based on some of these features and using the appropriate Rewriting Rules we hypothesize a likely identity for the surface under our feet. Then we use some or all of the other features to test and verify, or reject, or modify this hypothesis. This continuous process of hypothesising and testing (or variations there off) constitutes much of how knowledge can be used during understanding of images. We call the knowledge of activating the appropriate Feature Extraction Operators and Rewriting Rules to achieve this understanding, the Flow Of Control.

The hypothesize-test paradigm is provided here as an example of a kind of control flow. It is by no means the only one one should take into account. The three knowledge classes outlined above are provided merely to structure the problem area of knowledge used in Image Understanding into manageable subparts. They should not be seen as factors biasing our analytical findings.

A typical protocol is provided in Table 1, where the subject works with the stimulus in Figure 1. There are two kinds of evidence in the protocols: one, direct evidence represented by the subjects' thoughts about their own behavior in the entries entitled "CONJECTURE" and "DRAW"; two, the indirect evidence where a series of questions and answers have suggested to us certain behavioral patterns. A software (PROTDO) was developed to achieve consistency and objectivity in interpreting protocols. (Akin and Schultz, 1976) PROTDO is not a general purpose analysis program but rather an interactive filing system equipped with special search and format features tuned to the specific tasks of this investigation. Both categories of evidence from the protocols shall be discussed with respect to all three sources of knowledge outlined above, Feature Extraction Operators, Rewriting Rules and Flow of Control.

A. PERFORMANCE OF SUBJECTS

Before going into the details of knowledge acquisition it is useful to judge the level of understanding each subject achieved at the end of the sessions. All subjects were asked to describe the picture they were looking at in their own terms, after each session. Table 2 contains the complete descriptions provided by the subjects.

All subjects with the exception of S1, S2, and S5 were never shown the set of pictures the stimuli were selected from prior to the experiment. The other subjects were familiar with these images by virtue of their daily activities in the AI Laboratories, at Carnegie-Mellon University. However this does not present any experimental drawback because we are interested in getting at a wide range of knowledge applicable in Image Understanding rather than explain the problem solving behaviors of the subjects. The level of understanding of the subjects, with prior knowledge of the pictures, did not differ greatly from the two of the three remaining subjects at the end of the sessions anyway. Further more they had to work for it just as hard. Subject 2 who spend about twice the time on the task achieved superior understanding with respect to all subjects.

All subjects, except S3, had some accurate internal representation of the contents of the stimulus image. These accounted correctly for roughly 30 to 70 percent of all objects in the scene, depending on the particular subject and the time spent in the session. The reasons for S3's achieving a sub-standard level of understanding lie in the semantic misunderstandings that filled up a major portion of the 2 hour session.

On the other hand all 6 subjects had no difficulty in visually identifying the scene among 19 others once the session was over. Visual recognition occured in all instances based on very few general features found or lacking in the scene; S1: "not enough blue", S1: "no green at bottom." These features are based on low level information, i.e., color distributions, rather than high level concepts, i.e., car, building, etc. Consequently even S3 who had no idea what the scene contained had no trouble recognizing the scene after the session, based on a few low level features.

B. KNOWLEDGE ABOUT FEATURE EXTRACTION OPERATORS

First we scanned all protocols for physical components of the scene that were directly named by the subjects. Six levels of description have been referred to by the subjects; *scene, cluster, object, region, (sub-region,) segment*. In addition, other spatial and representational concepts have also been used; i.e., ***point, plane, space, 2-dimensional,*** and *3-dimensional*. (Table 3)

Secondly, all features refering to such components or their relations have been identified in the protocols. Seven such classes of features have been observed: *location, size, shape, quantity, color, texture, patterns* and *miscellaneous others*. A complete list of the Feature Extraction Operators is provided in Table 3.

Sixteen different classes of Feature Extractors were used to indicate locational relations. These were *delimiter, above, below, adjacent, around, along, far, within, without, center, corner, left, right, across, vertical* and *horizontal*. Some of these Operators were expressed in alternative wording; such as, *separator* for *delimiter*, *higher* for *above*, *surrounding* for *around* and so on. Similarly all Feature Extraction Operators discussed below represent classes with more than one alternate term in each class.

All common geometric shapes --i.e.,*square, circle, triangle, polygon, trapezoid*-- were used as shape Operators. In addition some not so common shapes were also used, such as, *t-junction, bifurcated*. Some other shape properties commonly used in the protocols dealt with linear elements and their combinations such as, *angularity, linearity, curved, flat, convex*.

Only five classes of Feature Extractors were used to specify size in the protocols. These were *large*, *small*, *long*, *short*, and *ratio*. This was largely due to the fact that the subjects were given information about the metrics of the various subparts of the scene, to the nearest 1/16 of an inch.

Many of the quantity Operators were imprecise concepts such as *some*, *most*, *more or less*, *few*, *extent*. However these did not pose difficulties in the interaction of the subjects and the experimenter. Other Operators of quantification were more precise in the sense that they were expressible in numbers or a clear criteria for evaluating them existed. These were *whole*, *any*, *quadrant*, *more*.

Color was the Feature Extraction Operator that was used most frequently in labeling regions. Most common hues were used extensively by the subjects. Also the density, contrast and texture of these hues were used to further specialize the coding schemes based on colors.

All of the Feature Extractors classified under *patterns* indicate some property of the relationship between multiple elements. Usually this property deals with the rate or nature of change of some feature between different subparts of the scene. For example, some of these pattern Operators are, *(in)homogeneity*, *gradual*, *abrupt*, *same*, *varying*, *continuous*, *(ir)regular*, *random*, *mixed*, *intersect and distribution*.

A set of commonly used Feature Extractors did not seem to fit readily in any of the above categories. These were categorized under *miscellaneous* and consisted of *approximate*, *relative*, *open*, *complex*, *basic* and *each*.

C. KNOWLEDGE ABOUT REWRITING RULES

All protocols contain many instances where information available in one level of scene description (feature, segment, region, object, cluster of objects, scene) is used to generate information in a different level. For example, a feature such as *green* in color, or a region *trapezoidal* in shape can be rewritten as *grass* or *building* in the object level, respectively. These elements of knowledge used in rewriting information available in one level of scene description into a different level are called Rewriting Rules (or Hypothesis Formation Rules).

Even though the protocols contain many instances where Rewriting Rules are used, none of these instances contain rules that are explicitly stated. For example, "Blueband is not a viewer or anything flat"; "Probably sky, if this is an outdoor scene"; "Maybe we have a road"; are some direct quotes from different protocols. All of these represent inferences made about the scene *using* the kinds of Rewriting Rules we are after. The existence of *blue* is used to infer *sky* in the above example, based on a rewriting rule such as "skies are usually blue".

Needless to say, these Rewriting Rules can only be *infered* from the evidence present in the protocols. The method we devised for identifying the Rewriting Rules consists of an interactive protocol transcription system. A program (PROTDO) was written to sort parts of the protocol into some predetermined categories and allow the analyzer to fill in other predetermined categories manually. The categories used consist of the *information* gathered from each answer given by the experimenter and the *inferences* made, based on the cumulation of information up to that point in time, as well as the *generation* and *testing* of hypotheses. Often in relating the information obtained to the inferences made the experimenter had to deduce the appropriate Rewriting Rules that were possibly used by the subjects in between the two.

Table 5 contains a sample transcription corresponding to the fist seven questions of the protocol in Table 1. The Rewriting Rules are labelled appropriately in Table 5. Notice the code provided in the parentheses after each rewiting rule. This code indicates the origin direction and destination of the inference enabled by that Rewriting Rule with the six scene description levels. For example, "Feature to Object" indicates that the rule rewrites information from the feature level into the object level.

Table 4 contains a complete listing of all Rewriting Rules observed in the transcribed protocols. Six other categories of use are identified for the Rewriting Rules. Some Rules are used as *assertions*, stating the existence of a descriptor at the destination level, while others are used as *negative assertions* refuting the existence of a descriptor. *Context-free rules are used more or less independent of prior information about the scene, while conditional rules contain a priori* conditions that must hold so that they can be applicable. Finally, *generative* rules are used to hypothesize and *analytical* rules are used to test these hypotheses.

1. Assertions

These are the assignments of certain descriptive terms, such as, red, big, car, grass, etc. to one or more components of the scene. Some examples are:

Green region is grass. (Feature to Object)

A blue region may possibly be the sky. (Region to Object)

2. Negative Assertions

These indicate that an assertion does not hold for the given components of the scene in question. At first this sort of information seems to be useless due to the many degrees of freedom there are in identifying the component being examined. However, negative assertions support hypotheses just like regular assertions, by negation. For example, the lack of a certain feature may support a hypothesis.

Sky and distant objects of similar color do not have
contrast edges. (Cluster of Objects to Feature)

3. Context-free Inference

Some inferences seem to depend on previous assertions and others do not. The latter are called context-free.

Perspective distorts shapes. (Scene to Region)

Grass has texture. (Object to Feature)

4. *Conditional Inferences*

The inferences that can be made only in the existence of certain assertions are called conditional inferences:

> Trapezoidal surfaces are the faces of rectilinear objects if they are in perspective. (Region to Object)
>
> A boundary if appropriately positioned with respect to a road may indicate that the road may have multiple lanes. (Scene to Cluster of Objects)

5. *Generative Inferences*

Inferences which are used to generate a hypothesis or an assertion are called generative. In the case of the picture puzzle paradigm all the information available to the subject consists of low level scene descriptors. Hence all hypothesis building based on this information works in a bottom-up fashion. That is information obtained in the low levels are used to hypothesize objects in the higher levels of scene description. For instance:

> Low contrast edges belong to very distant objects. (Segment to Object)
>
> Longitudinal lines on roads are the divisions indicating multiple lanes. (Segment to Object to Cluster of Objects)

6. *Analytical Inferences*

Inferences which are used to test an already generated hypothesis or an assertion are called analytical. By the token that generative inferences usually work bottom-up analytical inferences that test the hypotheses generated work in a top-down fashion. That is a hypothesis generated about a high level object is usually tested by verifying the existence of some low level properties of the object.

> Eyes, or eye-glasses, may look like two adjacent arcs. (Object to Segment)
>
> Man-made objects contain repetitive shapes. (Object to Region)

D. KNOWLEDGE ABOUT FLOW OF CONTROL

The behavior of all six subjects can be described as resembling the hypothesize and test paradigm very closely. Whatever the current focus of attention a subject has, he forms some hypothesis about what the property of one of the scene's contents is. Such as, "the scene is indoors"; or "there is a car in the scene"; or "the blue region is hilltops".

Then the subjects rewrite the hypothesis into a testable proposition by identifying some properties that may prove or refute the hypothesis. For example, using the above examples about the indoor scene, the car and the hill a typical subject would test the following propositions:

> indoor scenes may contain large wall areas that may be white or off-white in color.
>
> cars have shiny round accessories which occur in several locations.
>
> a sloping surface has its texture getting finer as it moves away from the observer.

These Rewriting Rules can be used for generating testable propositions as shown above or for generating hypotheses about the scene. Given the benefit of the answers to these tests; i.e., that there is a large white surface or shiny round regions or texture getting finer as you move up in a region the above hypotheses can all be generated respectively.

The third operation subjects seem to apply is related to the assessment of the progress they are making in performing the picture-puzzle task. Occasionally the issues at hand will be resolved, elaborated, or abandoned and the set of questions asked will exhibit a *shift* in the *attention* of the subject. For instance, obtaining the information that "there aren't sufficiently large white regions" in the above example (from the protocol in Table 1) leads the subject to revise his hypothesis about the "indoor scene." Later, the subject assumes that the scene is "outdoors with a sky", after having refuted the "indoors" hypothesis. Upon the rejection of this hypothesis he revised the "outdoor" hypothesis to "outdoor scene with an occluded sky".

The three operations; hypothesize, test, and shift of attention of search are iterated throughout all six protocols. Each of these operators appear in different forms as illustrated by the following examples.

1. Generate Hypothesis

Naming: After acquiring some information about the scene the subjects seemed to use free-association to name these entities as familiar objects. For example after discovering a large piece of grey area Subject 1 says "Maybe we have a road." The discovery of round eye-glass like objects prompts the assertion "People?" from the same Subject.

Backtracking to try a different hypothesis: If it was apparent after some examination that the current hypothesis was not supported by the evidence, the subjects proposed the opposite of the current hypothesis. For example if the subject was testing a hypothesis about the scene being an outdoor scene he would soon test whether it could be an indoor scene. This sequence is observed in Subject 1. Similarly after determining the orientation of a surface the same subject reverses his hypothesis about the object being a flat object and starts hypothesising about non-flat objects, "Blueband is not a river or anything else flat. Maybe a hill?"

In cases where a likely hypothesis and its exact opposite were tested and both were not supported, subjects proposed less probable but plausible hypotheses. For example in the following instance the subject consideres a special kind of outdoor scene after determining that the scene is neither an indoor nor an outdoor (in the normative sense, i.e., with a sky) scene.

> Neither outdoor nor typical office scene.
> How many regions are there in the scene. . .
> Is there a lot of green in the picture. . .
> Aha, maybe outdoors with blocked sky.

Sub-goal generation in the presence of uncertainty: In order to deepen the inquiry about a region it can be decomposed into smaller components, using uniformity of at least one property as the criteria of decomposition. More often the the criteria used to detect uniformity was "color":

> Is this the same color and texture throughout?

> In the region to the left of the biggest circle: what is the color and are there any areas significantly different in color.

2. *Test Hypotheses*

Exhaustiveness of Testing: When a hypothesis was to be tested the subjects inquired about all salient visual properties of the hypothesized component, *exhaustively*. Most subjects start out identifying a particular region by asking about its color, shape or location with respect to a known region. After identifying a region, it takes on the average 3 to 4 more properties to identify before that region can be successfully incorporated in the total understanding of the scene. Hence a total of 4 to 7 properties are explored about each region. And this is almost exhaustive of all classes of Feature Extraction Operators used by the subjects.

Salient feature testing: The principle of exclusiveness is violated under certain conditions. If the hypothesized property can be adequately represented by one major salient feature, the presence/absence of that feature could be decisive in accepting or rejecting that hypothesis. For example consider the following conjectures made by the subjects:

> *trapezoids* is looking for perspective line. .

> *Emptiness* was looking for maybe number of areas, indicating number of objects.

Whole and its Parts: Most entities in the scene are parts of other "things" while they are made up of smaller parts too. Usually parts and wholes are related to each other at least along one Feature Extraction dimension. Salient features of entities can be used to associate spatially unrelated regions as parts of the same object or object cluster. This can be done in one of two ways:

One, initially unrelated regions may be parts of the same object. Consider the following examples:

> Sudden thought: gray lines are border of the gray bottom objects?
>
> Are the two reddish brown areas (separated by the
> tannish white area) connected?

Two, there may be undiscovered regions in the scene which are parts of the object currently being examined. As shown by the following example:

> Maybe this is the front of a car: headlights, etc.
> Let's try looking for some circles.

3. *Focus of Attention in Search*

The most powerful tool of the subjects seems to be the ability to deal with incomplete and erroneous hypotheses. Given any arbitrarily likelihood of success for a hypothesis, the subjects can operate either under the assumption that it is, or is not, true until in fact it is eventually confirmed one way or the other. The degree of confidence associated with an assumption seems to be irrelevant to the usefulness of that hypothesis due its tentative nature. Given a state of information about the scene the subjects have the options of pursueing, abandoning, accepting, modifying or striking the current hypothesis.

Discovering the gist of a scene: The first hypothesis each subject deals with has to do with the issue of the "gist" of the scene. All first five subjects ask their first questions about the context or gist of the scene.

> Are there colors in the scene?
>
> Are there some high contrast edges in the middle of the scene?
>
> Does the picture contain wide open space?
>
> I assume the picture is representational?
>
> Is the photo square or rectangular?

The only exception to this is the sixth subject. He starts by dividing up the scene into quadrants and then asks about the general line, texture and intensity content of each quadrant. This is the only truely bottom-up approach we observed in our experiments.

Use of salient feature in the solution of next issue: The selection of the next issue or hypothesis to inquire about is another vital aspect of Flow of Control. Usually the next issue selected is based on a dominant descriptor. Subjects inquire about the largest region or the one with most contrast edge first. The following examples illustrate the use of size and contrast in determining the significance of a given region's property.

> Any red in the picture?
> Experimenter: Yes, two reddish areas of very small size.
> Forget it. (small size is not important)
>
> Describe the location and approximate shape of largest homogeneous region.
>
> Are there some high contrast edges in the middle of the scene?

On the other hand, if an allogether new issue is necessary, a dominant property based on the current gist of the scene, such as locational adjacency, etc., can be used to explore new regions. All three examples below illustrate the use of adjacency in selecting the next region for examination.

> Concentrate next on the lower edge of the sky.
>
> I'll try working from the boarder inward.
>
> Work by process of elimination from the edges.

When there was no guidance from the gist assumptions, the subjects went back to unresolved issues. For example, Subject 1 says "Look at the supposed road," after dealing with another issue for a while; similarly Subject 2 notes "I don't feel any need to continue at this point in the lower region of the scene. . . .back to get more detail on blueband and green region."

Similarly, when a new piece of evidence emerges from an inquiry and it cannot be simply accommodate by the current assumptions about the scene, this issue can be set aside for future exploration. Subject 2 notes "Interesting, but come back to this later. "

Resolution of hypothesis: Normally all hypotheses get resolved after a number of repeated inquiries about it. Sometimes it takes to come back to an issue after other issues have been resolved. This is inevitable due to the conditionality of Rewriting Rules in general. For example, a blue region in a "sea" scene is likely to be the sea and/or the sky, while in a "portrait" scene it is likely to be a piece of garment or a background surface. And these issues can be resolved after determining the context of the scene and relative location of these regions with respect to others.

On the other hand some hypotheses can not be resolved by using the most probable associations. In such cases some rarely used Rewriting Rules are used to justify some of the findings in spite of some apparent contradictions. Consider the following examples:

> I'm beginning to think this whole thing is a Kandinsky painting.
>
> Almost certainly a city scene. Still puzzled by the low contrast bottom edge of blueband, the green region within blueband, and identity of blueband. Perhaps it's sky also and its different color is because of pollution of sunset.
>
> Don't know what they are? Blue clouds or clouds which look blue because of lighting.

IV. CONCLUSIONS

This study attempts to specify knowledge used in an image understanding task by human subjects. In order to achieve this, a picture-puzzle paradigm has been developed.

The first type of evidence provided by the protocols is the knowledge about possible primitive Feature Extraction Operators. The range of operators seem to be modest, yet we suspect this was caused by intrinsic properties of the picture-puzzle paradigm. Translation of visual information into the verbal domain may have taken away from the richness of the visual information. Yet we believe that the operators represent a desirable subset in any system for computer understanding of scenes.

The second type of evidence, i.e., knowledge about the Rewriting Rules found in the protocols seems natural and appears to be easily implementable through production system-like schemes. While this set of rules is not intended to be complete and exhaustive they provide a good beginning for analysis.

The Flow of Control found in the protocols reveal some general techniques that already appear to be useful in "blackboard" model-like schemes. (Erman and Lesser, 1976) Further experience from our laboratory indicates that different tasks used with the picture-puzzle paradigm require different Flow of Control mechanisms. We recommend special attention to task properties in all studies so that an optimal task specific control strategy might be utilized.

V. ACKNOWLEDGEMENTS

We wish to thank Ron Ohlander and Keith Price for the many hours of work spent on the experimental tool and Herbert Simon, John Gaschnig, Adrian Baer, Bruce Lowerre, Randy Korman, and John Kender for the valuable protocols they provided.

REFERENCES

1. O. Akin and M. Schultz, An Interactive Protocol Analysis System for Knowledge Acquisition in Image Understanding, Computer Science Department Reports, Carnegie-Mellon University, 1976.

2. G. W. Baylor, A Treatise on the Mind's Eye: an Empirical Investigation of Visual Mental Imagery. Ph.D. Thesis, Carnegie-Mellon University, 1971.

3. G. T. Buswell, *How People Look at Pictures*, University Press, Chicago, 1935.

4. L. D. Erman and V. R. Lesser, A multi-level organization for problem solving using many, diverse, cooperating sources of knowledge. Working Papers In Speech Recognition IV, Carnegie-Mellon University, February 1976.

5. A. M. Farley. VIPS: A Visual Imagery and Perception System: the Results of a Protocol Analysis, 2 volumes. Ph.D. Thesis, Carnegie-Mellon University, 1974.

6. O. Furschein and M. A. Fischler, A study in descriptive representation of pictorial data, *Pattern Recognition*, 4, 1972.

7. A. Guzman, Computer Recognition of Three-Dimensional Objects in a visual scene, MAC-TR-59, Ph.D. Thesis, MIT Project MAC, 1968.

8. J. Hochberg, *Perception*. Prentice Hall, New Jersey, 1964.

9. J. Hochberg, In the mind's eye, *Contemporary Theory and Research in Visual Perception* (R. N. Haber, Ed.) Holt, New York, 1968.

10. B. Julesz,*Foundations of Cyclopean Perception*, University Press, Chicago, 1971.

11. M. D. Kelly, Visual Identification of People by Computer, AIM-130, Ph.D. Thesis, Stanford University, 1970.

12. G. R. Loftus, A framework for a theory of picture recognition, presented at the National Academy of Sciences Specialist's Meeting on Eye Movements and Psychological Processes, Princeton, New Jersey, April 1974.

13. T. P. Moran, The symbolic nature of visual imagery, Proceedings of Third International Joint Conference on Artificial Intelligence, Stanford, California, 1973.

14. A. Newell, On the analysis of human problem solving protocols, *Calcul et Formalisation dans les Sciences de L'Homme* ed. by J. C. Gardin and B. Jaulin, 146-185, Centre National de la Recherche Scientifique, Paris, 1968.

15. R. B. Ohlander, Analysis of Natural Scenes, Ph.D. Thesis, Carnegie-Mellon University, April 1975.

16. R. B. Ohlander, R. Reddy and O. Akin, An Experimental System for Knowledge Acquisition in Image Understanding, Computer Science Department Reports, Carnegie-Mellon University, 1976.

17. R. Reddy and A. Newell, Image Understanding: potential research approaches, Unpublished report, Computer Science, Carnegie-Mellon University, 1975.

18. R. N. Shepard, Recognition memory for words, sentences and pictures, *Journal of Verbal Learning and Verbal Behavior*, 6, 1976.

19. D. L. Waltz, Generating Semantic Descriptions from Drawing of Scenes with Shadows, AI TR-271, MIT-AI Laboratory, November 1972.

20. W. A. Woods and J. Makhoul, Mechanical inference problems in continuous speech understanding, *Artificial Intelligence* 5, 1974.

TABLE 1 Sample Protocol from Task 1.

EXP: BEGIN

SUB: are there colors in the picture?
EXP: yes

SUB: is the color blue predominant at the top of the picture?
EXP: no

SUB: is the color white predominant throughout?
EXP: no
CONJECTURE: neither outdoor nor typical office scene???

SUB: how many regions are there in the picture?
NOTE: is this subject familiar with the set of stimuli.
EXP: there are many regions and it is difficult to determine regional boundaries

SUB: is there a lot of green in the picture?
EXP: yes
CONJECTURE: aha. maybe outdoors with sky blocked

SUB: are there 'any pronounced straight lines in the picture?
EXP: yes
CONJECTURE: buildings?

SUB: do the lines form angles with each other that are about equal in degrees?
EXP: no
CONJECTURE: buildings?
CONJECTURE: try subdividing the picture

SUB: is there any consistent color or texture in the upper left corner.
EXP: yes,green

SUB: is this the same color or texture throughout?
EXP: it varies slightly
CONJECTURE: big help!

SUB: is there any brown in the picture?
EXP: yes there are very small brown areas, sort of little patches
DREW: Feature #1

SUB: what color are the lines?
EXP: one is black, another is white a third one not very even is gray
DREW: Feature #2
CONJECTURE: lets find out about orientation of
the lines. they might be telephone wires?

SUB: what is the orientation of the black line with respect to the bottom? (angle)
EXP: it is approximately 10 degrees with the bottom edge of the picture, it
however does not touch the bottom line

SUB: does this line extend left to right across the entire picture?
EXP: not across the entire picture, left to right at an angle, yes

SUB: is the picture longer than it is tall?
EXP: yes

SUB: approximately ratio of length to height?
EXP: 2/3
DREW: Feature #3

SUB: is the black line near the bottom of the picture?
EXP: no

SUB: about how close to the bottom, relative to the picture height?
EXP: it is at a central location in the picture with respect to the picture
frame it is slightly to the left than the dead center
DREW: Feature #4

SUB: where is the white line?
EXP: it is below the black line and almost parallel to it
DREW: Feature #5
CONJECTURE: maybe we have a road?

SUB: is the color or texture between the white and black lines more or less
uniform
EXP: yes, gray
CONJECTURE: looks good on the road
DREW: Feature #6

SUB: is the gray line part of this gray region?
EXP: no
CONJECTURE: hmm. maybe it was an overcast day. nope: mostly green.

try it anyway.
SUB: is the top part of the picture:does it have a gray region?
EXP: no
CONJECTURE: look at the supposed road
DREW: Feature #7

SUB: the gray area bordered by the black and white lines: what percentage of
the picture is it?
EXP: 3-5 percent

TABLE 2. SUBJECTS' UNDERSTANDING OF IMAGES.

S1: Car Scene (2:30)*

It is outdoors. There appears to be a road with a car. [Can you identify other things or objects other than the ones you mentioned?] There are other objects: I guess they are man-made. If they are on the car, they are headlights or other mechanical parts. There is a lot of grass and some ground.

S2: Downtown Scene (4:35)

There is a blue sky at the top of the picture, bordering the ridge of a hill, beyond which on the right one can see the ridges of two more distant hills. The main hill is green-gray-bluish colored, turning to less blue and more gray, green and brownish-red as one goes down the hill. On the hill is a more greenish region within which are 5 or 6 thin horizontal short strips and some light-colored spots. I conjecture that they are buildings and other artifacts, might be a housing development. In the foreground is a city scape, probably downtown with many tall buildings. The buildings occupy more than 60 percent of this lower portion of the scene, much of the rest of the lower (foreground) consists of a pond of water on the left near the largest of the buildings and another pond in the center bottom [wrong] of the scene, and a strip of green, probably just grass extending left to the center of the scene in the lower portion of the city scape.

S3: Office Scene (2:35)

[No high-level concepts formed.]

S4: House Scene (2:45)

The horizontal streak with white lines is a street or road. The road passes in front of some buildings -- or billboards. I think it is a landscape with blue sky above and mainly green grass and shrubs below. There is a bush of some size in the lower left and a tree in the foreground to the right of center. A street or road across the scene horizontally. I have no good hypotheses about the nature of the small rectangles (since they are flat, not solid). The vertical objects with rounded tops could be silos. I have no idea what the wedge-shaped objects are.

S5: Bear Scene (2:15)

Facts I know for sure: 1) There is a very large dark area in the center of the picture. 2) This area is roughly pear (or bell) shaped and seems to have a bit of the area extending lower than the rest. 3) The background contains many brown, gray, and tan areas that are confusing. 4) There are some lighter areas within the central dark region: two white areas, one in the center and the other near the right top. There is also a cluster of smaller patches in the central lower portion of the dark area.

Things I think I know: 1) It is a picture of a bear sitting upright with his right hind leg folded. 2) I believe he is facing to his left (the white area may be his nose), but I'm not sure. 3) The confusing background is rocks (from a zoo).

S6: Portrait (2:50)

The image appears to be that of a man sitting or standing erect and facing frontally. Behind him is an undifferentiated field of gray/white. The man seems to have much hair including a beard, and may be wearing eye-glasses. Another possibility is that it is a women with her hair partially draped over her face or possibly it is a picture of an ape. But small arcs suggest otherwise. Finally, large arch is suggestive of some clothing artifact or subject is possibly holding some object in front.

*Total time (in hours) the subject was allowed to work on the session.

TABLE 3. FEATURE EXTRACTION OPERATORS

1. COMPONENTS OF SCENES: things, elements

scene	picture, content
cluster	
object	
region	figure, clouds, area, patch
sub region	sub figure, spots, part, sub area, part of
edge	line, trace, segment, side, boundary, border, contour, base, outline
point	apex
plane	surfaces
space	

2. LOCATION

Delineator	interrupting, between, separator
above	upper, top, high(er), over
below	low(er), down, bottom, under
adjacent	connect(ed), come close, touch, close(r), meet at, attached, near, joined, far, beyond
without	outside, non overlapping, outer
within	in, part of, overlapping, inward, throughout, interior, through
center	middle
corner	last
left	
right	
across	
around	bounds, boundary, border, surrounding
along	parallel
vertical	top to bottom
horizontal	left to right
contain	enclose, bound, bounded, contained, part of

3. SHAPE

rectangle	
square	
circle	circular
ellipse	
parallelogram	
trapezoid	
triangle	wedge
elongated	band
angle	sharp, jagged, wedge
linear	straight
curved	
axes	
flat	
T junction	
convex	
bifurcated	
polygon	

4. SIZE: dimension, length, height, distance

large	big
small	little
long	thick, wide, tall
short	narrow,
ratio	percentage

5. QUANTITY

some	portion, partial
most	lot, good deal, lots of, many
whole	entire, all, complete(ly)
half	between
quadrant	
any	
more	further
less	
few,	
more or less	
extent	degrees, grade
ratio	percentage, density
high	
low	
just	nearly
predominately	generally, dominately, overall, throughout, significant, common
hardly	
quite	
average	

6. COLOR hue, tone, density

contrast	
white	whitish, tannish white
green	
red	
yellow	
blue	greenish blue
brown	brownish, reddish brown
clear	
black	
gray	grayish
silver	
pink	
shades	
light	clean, pale
dark	

7. TEXTURE:

checkerboard	
smooth	homogeneous, uniform, even(ly)

8. PATTERNS: order

homogeneous	uniform, smooth
unhomogeneous	distorted
gradual	blending
abrupt	sharp
varying	alternating, change, distorted
same	equal, similar
continuous	continue
regular	repeated, consistent
irregular	distorted
random	
mixed	interrupt, intervals, isolate
distribution	
different	other, various, vary, besides
fuzzy	blending, distorted, indefinite
discernible	determinate, pronounced, unobstructed, bright, sharp, clear

9. REPRESENTATIONAL

2 D
3 D
perspective(ly)

10. OTHER QUALIFIERS: features

approximate	about, nearly
relative	with respect to
open	
complex	
basic	primary, prime
each	

TABLE 4. REWRITING RULES.

Region to Feature

1. Surfaces of roads are gray.

Object to Feature

1. Sky is blue.
2. Soil is brown.
3. Walls are usually white.
4. Roads have uniform color and texture.
5. Tires have a definite range of colors.
6. Clouds are white.
7. Surfaces sloping up and away from observer -- hills, etc. -- are closer at the bottom than at the top and higher (3-D) at the top than at the bottom.
8. Grass is flat in 3-D.
9. Grass is textured.
10. Cars have silver colored accessories.

Cluster of Objects to Feature

1. Distant objects look like a heap of in-homogeneous and non-uniform colors.
2. Greenery is green.
3. Human's complexion is pink or pale pink.

Segment to Segment

1. Man made shapes are usually repeated more than once in a scene.
2. Segments broken into discontinuous parts by occluding objects are co-linear.

Region to Segment

1. Surfaces are defined by edges.
2. Parts of an occluded region separated by other objects occluding it are the same color(s).
3. Trapezoids have two opposite edges that intersect when extended.

Object to Segment

1. Eyes, or eye-glasses, look like two adjacent arcs.
2. Very distant objects have low contrast edges.
3. Telephone wires usually run horizontally.
4. Buildings have edges.
5. Building facades have checkerboard texture.
6. A water-body is a greyish-blue flat surface.
7. Objects in perspective have edges, or lines, vanishing to a common point.
8. Corners of rectilinear objects in perspective are made up of edges forming I-junctions.
9. Man-made objects are bounded by straight edges.
10. Objetts in the scene usually occlude the horizon line.

Cluster of Objects to Segment

1. Sky and distant objects of similar color do not have contrast edges.
2. Clusters of vertical objects have predominantly vertical edges, or lines.
3. Sky and distant landscape form curvilinear edges.
4. Sky and man made objects form jagged edges.
5. Multiple lane roads are divided by boundaries, or lines.
6. Buildings have lots of vertical edges, or lines.

Segment to Region

1. Lines, or edges, define surfaces.

Region to Region

1. L-shaped surfaces, or their relations are quadrilaterals occluded by another quadrilateral.
2. A horizontal rectilinear surface in perspective looks trapezoidal.
3. A trapezoidal shape may be a rectilinear surface in perspective.
4. A region occluded by all others is the most distant region, like the sky region,

Object to Region

1. Sky is above.
2. Walls may occupy a large section of a scene.
3. Rectilinear objects have trapezoidal faces in perspective.
4. A tree is an object consisting of elongated brownish vertical rectangle, to the top of which is attached a more or less convex green mass of indeterminate shape.
5. Car headlights are usually silvery adjacent and, circular.
6. Most buildings have rectangular surfaces.
7. Rectilinear objects are made up of straight line quadrilaterals which share sides.
8. Unoccluded surfaces of rectilinear objects are trapezoidal in the 2-dimensional representation.
9. Trees have vertical brown trunks.

Cluster of Objects to Region

1. Rectangular objects occluding each other have straight line polygonal surfaces with adjacent edges.

Scene to Region

1. Perspective distorts shapes.
2. Outdoor scenes usually contain some fuzzy boardered regions with blending colors.
3. City scapes contain clusters of vertically oriented rectangular shapes.

TABLE 4. REWRITING RULES.
Continued.

Feature to Object

1. Green is grass.
2. Reddish brown is the color of hair.

Segment to Object

1. Contrast edges stipulate object boundaries.
2. Two adjacent arcs may be eyes, or eye-glasses.
3. A checkerboard texture is probably a building facade.
4. Edges, or lines, vanishing to a common point are the edges of an object in perspective.
5 T-junctions at corners of rectangular surfaces are corners of rectilinear objects.
6 Low contrast edges may belong to distant objects.
7. Straight edges belong to man made objects.

Region to Object

1. A blue region may possibly be the sky.
2. A brown region is possibly the soil.
3 White surfaces may possibly be walls.
4 White surfaces may possibly be clouds.
5. A surface closer to the observer at the bottom than at the top is an upward and away sloping surface -- i.e., hills, etc.
6. Silver colored parts, or accessories, may belong to cars.
7. Repeated shapes are probably man made objects.
8. A segment at the top of a scene could be the sky.
9. A large uniform surface in an indoor scene may bee the walls.
10. Trapezoidal surfaces are the faces of rectilinear objects in perspective.
11. An object consisting of elongated brownish vertical rectangle, to the top of which is attached a more or less convex green mass of indeterminate shape is a tree.
12 A pair of circular adjacent, silvery objects, may be the headlights of a car.
13. Rectangular surfaces usually belong to buildings.
14. Straight line quadrilaterals which share sides are parts of rectilinear objects.
15. Trapezoidal surfaces are the unoccluded surfaces of rectilinear objects.
16. Vertical, brown regions may be tree trunks.
17. Rectangular shapes in hair may be hair-pins or eye-glass frames.

Object to Object

1. Roads have cracks.
2. An object occluding another is closer to the observer.
3. Car wheels are spatially lower than the car body.
4. Buildings are located below the region.
5. Bushes are close to ground.
6. Shadows to objects touch bases of objects they are cast by.
7. Man made shapes are usually repeated in scenes.
8. Tires are around hub-caps.
9. Hair close to cheeks is beard.

Cluster of Object to Object

1. Sky has clouds.

Scene to Object

1. Some outdoor scenes may contain objects like the sea or sky (i.e., large scale) in the lower half of the visual field as well as the upper half.
2. Outdoor scenes may contain skies.
3. Indoor scenes may contain walls.

Feature to Cluster of Objects

1. Green is greenery.
2. Pink or pale pink is human complexion.
3. Regions with inhomogeneous color and texture may be indicative of a conglomeration of distant objects.

Segment to Cluster of Object

1. Sky and distant objects do not form contrast edges.
2. A boundary, if appropriately positioned with respect to a road, may indicate that the road is multiple lane.
3. Lots of vertical lines, or edges, may indicate buildings.
4. Many vertical lines, edges are indicative of clusters of vertically oriented objects.

Object to Cluster of Objects

1. The objects closer to the observer are the objects occluding the more distant ones.
2. Roads are closer to the bottom of one's visual field.
3. Eyes may belong to human beings.
4. Beard is hair near the cheeks.

Scene to Cluster of Objects

1. Things covering entire width of scenes are composite objects such as landscape, sea, sky, etc.
2. Outdoor scenes may have lots of greenery.
3. Outdoor scenes may have lots of buildings.
4. Large scale objects -- landscape, hills, sky, sea, etc. -- span usually the entire width of a visual scene.
5. All objects in outdoor scenes occlude the sky.

Region to Scene

1. Vertically oriented rectangular clusters are contained in city scapes.

TABLE 5. SAMPLE FROM TRANSCRIBED PROTOCOL OF SUBJECT 1.

1. SEARCH: The nature of the representation of the picture.
R.R.[1]: There are colors in real scenes. (feature to scene)
R.R.: Photographs may be colored or black-and-white. (scene to feature)
HYPOTHESIS: There are colors in the photograph.
F.E.O[2]: COLORS
TEST: 'ARE THERE COLORS IN THE PICTURE?'
[ANSWER: 'YES']
F.E.O: COLORS
R.R.: Picture is colored.

2. SEARCH: The nature of the contents of the picture.
R.R.: Scenes may be categorized into two: indoor and outdoor. (scene to scene)
HYPOTHESIS: The picture is outdoors.
R.R.: The pictures of outdoor scenes may contain the sky. (scene to object)
R.R.: The sky is blue. (cluster to feature)
R.R.: The sky is at the top. (cluster to region)
TEST: 'IS THE COLOR BLUE PREDOMINANT AT THE TOP OF THE PICTURE?'
[ANSWER: 'NO']
F.E.O: BLUE / AT TOP / DOMINANT.
R.R.: Picture is not predominately blue at top.
R.R.: There is no sky visible in the scene.
R.R.: The picture is not of an outdoor scene.

3. HYPOTHESIS: The picture is indoors
R.R.: The indoor-scene has predominately white walls. (scene to region)
TEST: 'IS THE COLOR WHITE PREDOMINANT THROUGHOUT?'
[ANSWER: 'NO']
F.E.O: WHITE / PREDOMINANT / THROUGHOUT
R.R.: Picture not predominantly white.
R.R.: There is no wall visible in the scene.
R.R.: The picture is not a typical office scene.
R.R.: 'NEITHER OUTDOOR NOR TYPICAL OFFICE SCENE'

4. SEARCH: Contents of the picture: 'NEITHER OUTDOOR NOR TYPICAL OFFICE SCENE'
R.R.: A scene contains objects. (scene to object)
R.R.: Objects are made up of a finite number of homogeneous areas (=regions). (object to region)
HYPOTHESIS: There are n number of such regions in the scene.
F.E.O: REGIONS
TEST: 'HOW MANY REGIONS ARE THERE IN THE PICTURE?'
[ANSWER: 'THERE ARE MANY REGIONS AND IT IS DIFFICULT TO DETERMINE REGIONAL BOUNDARIES']
F.E.O: HOW MANY / REGIONS
R.R.: There are many regions in the scene.

5. SEARCH: Identify one such region.
R.R.: Outdoor scenes may have lots of greenery. (scene to cluster)
R.R.: Greenery is green. (cluster to feature)
HYPOTHESIS: There is lots of greenery in the scene.
F.E.O: LOT OF / GREEN
TEST: 'IS THERE A LOT OF GREEN IN THE PICTURE?'
[ANSWER: 'YES']
F.E.O: LOT OF / GREEN
R.R.: The scene contains lots of greenery.
R.R.: Lots of greenery may block the sky.
R.R.: This may account for why no sky was found earlier.
R.R.: 'AHA MAYBE OUTDOORS WITH SKY BLOCKED'

6. SEARCH: The contents of the outdoor scene.
R.R.: Outdoor scenes may be scenes of man-made "scapes", i.e., townscapes. (scene to scene)
HYPOTHESIS: There are man-made objects in the scene.
R.R.: Man-made objects have straight edges. (object to segment)
TEST: 'ARE THERE ANY PRONOUNCED STRAIGHT LINES IN THE PICTURE?'
[ANSWER: 'YES']
F.E.O: PRONOUNCED / STRAIGHT / ANY / LINES
R.R.: There are objects in the scene with straight edges.
R.R.: There are man-made objects in the scene.
R.R.: 'BUILDINGS '

7. SEARCH: Identify some man-made objects.
R.R.: Buildings are man-made objects. (object to object)
R.R.: Buildings have straight edges. (object to segment)
HYPOTHESIS: There are buildings in the scene.
R.R.: Edges of buildings meet in angles that are about equal. (object to segment)
F.E.O: LINES / ANGLES / EQUAL / DEGREES
TEST: 'DO THE LINES FORM ANGLES WITH EACH OTHER THAT ARE ABOUT EQUAL IN DEGREES?'
[ANSWER: 'NO']
F.E.O: LINES / ANGLES / EQUAL / DEGREES
R.R.: There are no angles formed by the lines, that are about equal.
R.R.: The straight edges are not parts of buildings.
R.R.: There may not be any buildings in the scene.
R.R.: 'BUILDINGS?'

[1] Rewriting Rules.
[2] Feature Extraction Operators.

Figure 1. Car Scene

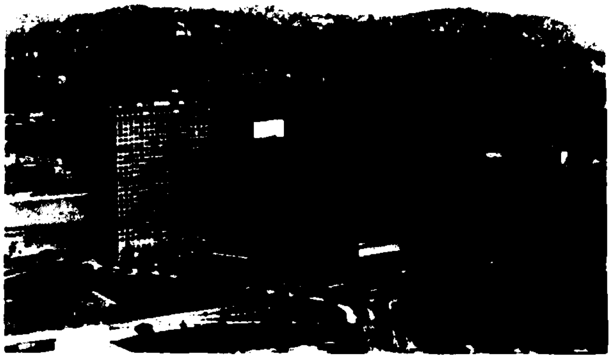


Figure 2. Downtown Scene

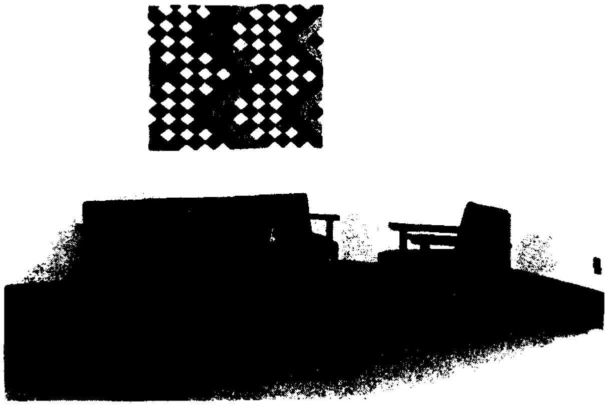


Figure 3. Office Scene

Figure 4. House Scene

Figure 5. Bear Scene

Figure 6. Girl Scene

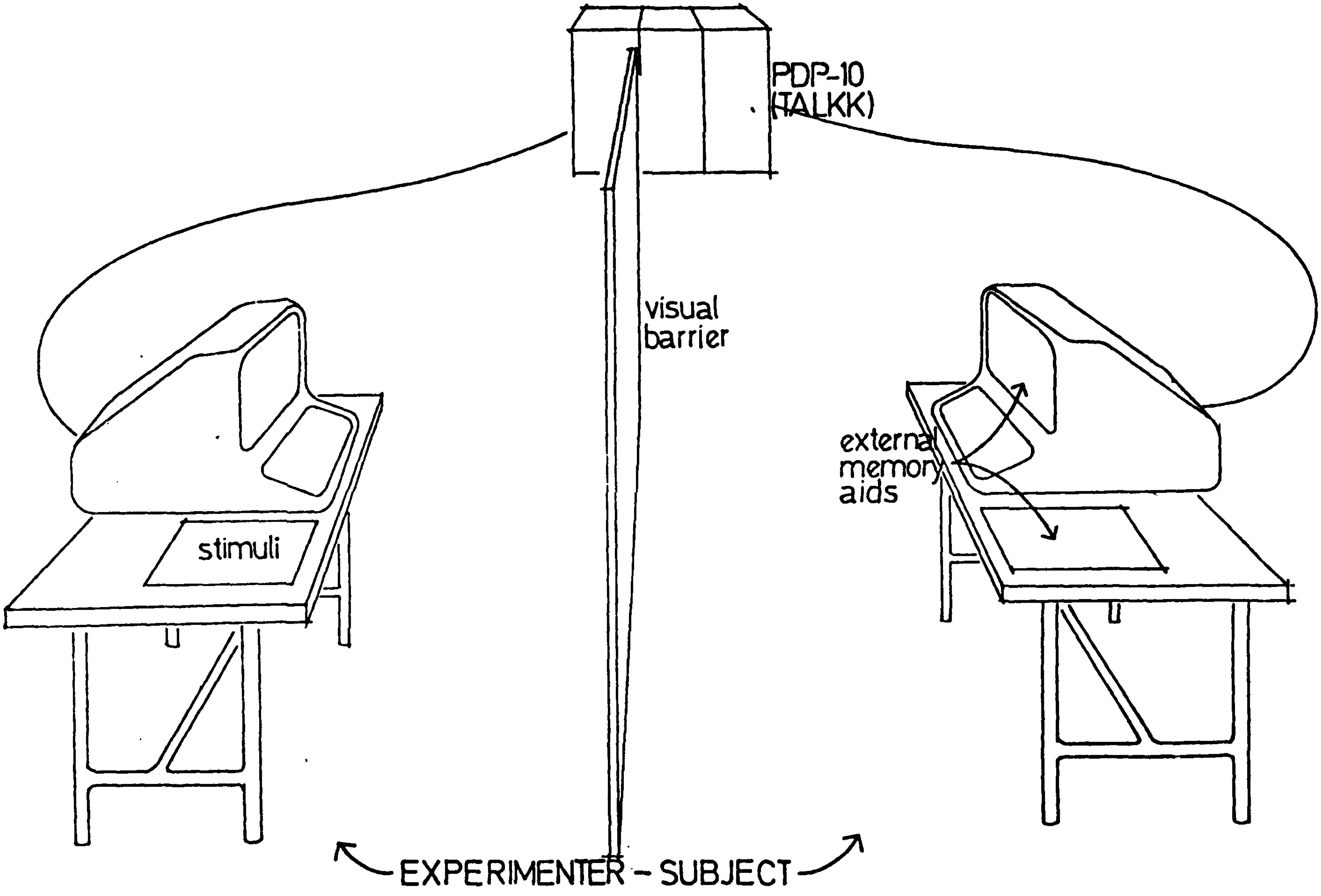


Figure 7. Experimental Set-up.

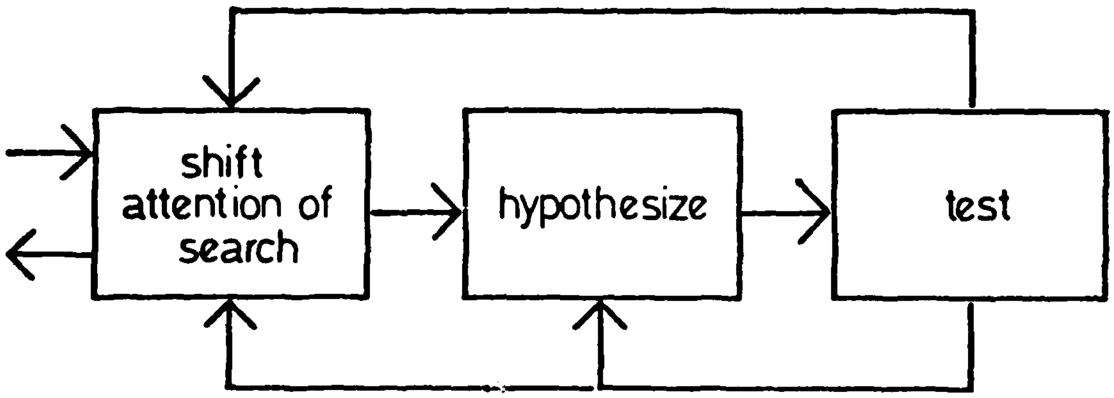


Figure 8. Phases of Control in the picture-puzzle task

An Experimental System for Knowledge Acquisition in Image Understanding Research

R. Ohlander, R. Reddy and O. Akin*
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

A class of problems related to model building in cognitive psychology and artificial intelligence require the codification of knowledge sources. Some cognitive tasks of interest which fall in this category are perception of speech, perception of visual scenes, or perception of other symbolic media such as maps, drawings, and written text.

Various experimental tools that have been developed until now can be categorized into three classes: eye fixation studies, protocol analysis of mental imagery related tasks, and protocol analysis with controlled exposure of stimuli.

Eye fixation studies have yielded specific information on the feature selection processes in perception. (Buswell, 1935; Loftus, 1974; Mackworth and Morandi, 1976) However, due to a lack of theoretical models of the image understanding process, these studies have not led to a codification of knowledge sources used in picture processing.

Protocol analysis studies of tasks with imagined visual objects provide theoretical models of image processing. (Baylor, 1971; Moran, 1973) However, like protocols based on limited exposure these do not provide direct evidence on the knowledge sources used in these tasks. (Farley, 1974; Potter and Levy, 1969) Basically, this is true because these tasks were not designed to explore sources of knowledge.

In this paper, we propose an experimental paradigm which is designed to explore the knowledge sources used in visual understanding tasks. Our *a priori* taxonomy for knowledge needed in image understanding is made up of three parts: Feature Extraction Operators, Rewriting Rules, and Elements of Control Flow. Feature Extraction Operations are based on visual properties found in scenes, such as color, shape, location, size, quantity, texture, etc. that can be used to decompose a scene into sub-parts and then label and characterize these sub-parts.

Rewriting Rules enable the translation of these low-level attributes into meaningful visual components -- grass, chair, table, room, etc. -- and vice versa. These components can be expressed as elements of various, hierarchical levels of scene description. For example, the color "green" when supplied as a low-level information, may help to infer a "leaf" at a higher level, or a "forest" at a yet higher level. The flow of control governs the use of Feature Extraction Operators and Rewriting Rules in the context of a specific goal-directed visual task. Elements of

* Also in the Department of Architecture.

Control Flow are helpful to develop alternative scene descriptions and/or test such descriptions in order to generate a final, unique description of the scene.

The "picture-puzzle" paradigm we developed aims to provide direct evidence for all three classes of knowledge cited above. Further, it provides a simulation of the process of machine understanding of visual scenes. The task of the subjects is to describe the scene including the parts of the visual scene, based solely on verbal question-answer interactions with the experimenter. The experimenter can answer questions concerning lower levels of scene description, only. For example, he is allowed to say that there is a "green region" with certain texture, size, location, shape, etc. However, he is not allowed to say that there is "grass" in the scene.

In conventional experimental conditions where subjects interact directly with a visual scene or image, the inferences made during the analysis of the data are either based on unobtrusive recordings of subjects behavior (eye fixations, reaction times) or the introspections of subjects about their own behavior during the task (protocols). Figure 1 represents the flow of information in the conventional case. Eye fixation and reaction-time information provides very little in terms of knowledge used. A major problem with protocols of self-assessment is the loss of much of what is internally processed.

Ideally, the experimenter needs to have first hand experience in monitoring or observing the interactions of the subjects with the stimulus. The picture-puzzle paradigm achieves the monitoring of the interaction adequately. Figure 2 indicates the schematic interaction between the subject, stimulus and the experimenter. All interactions between the subject and the stimulus go through the experimenter in the case of the picture-puzzle paradigm.

I. APPLICATIONS

Two video-terminals were used in the experiments. The terminals were connected to each other by means of software (TALKK program) to enable typed-in communication between their users. The facilities of the Computer Sciences Department at Carnegie-Mellon University were used to accommodate this set-up. The terminals were located such that no visual communication between their users was possible (Figure 7 of the first paper in this volume, entitled "Knowledge Acquisition"). The subject and the experimenter were able to communicate only verbally thru the TALKK program.

TALKK was designed to record all statements made by both the subjects and the experimenter throughout the experiments. It also enabled them to input conjectures and notes about the task at any time during the experiment. It further enabled subjects to correlate their personal notes and drawings, which they were allowed to make on separate sheets of paper, with the typed in protocol recorded by the TALKK program.

The stimuli used were produced for and used in automated image understanding research by Ohlander. (1975) All the scenes were constructed or selected as usual natural images of very familiar objects. All the stimuli in the first six figures of the first paper in this volume have been used in this experiment.

The subjects were simply instructed to "understand" the contents of the stimulus so that they would be able to describe all major objects in the scene. The experiments were terminated when the subjects thought they understood all major objects or at the end of 2 and 1/2 hours, which ever came first. The subjects were required to perform the experimental task with any one of the six different stimulus scenes.

Due to the fact that the experimental paradigm used here is totally novel, at least to our knowledge, it deserves a careful reconstruction of its proceedings for clarity. We suggest that the reader go over the sample protocol (in Table 1 of the first paper in this volume, entitled "Knowledge Acquisition") in which the subject tries to "understand" the given image (in Figure 1 of the same paper). Note that the "DREW #"s in the protocol refer to the personal notes of the subject indicated by numbers in Figure 5.

Since it was one of the independent variables being examined, the range of operators used in inquiries by the subject were not limited. However, when the subjects used high level descriptors (which were defined as illegal questions at the onset of the experiments) to inquire about the scene, the experimenter refused to understand the question, this forced the subjects to reformulate their questions causing them to use low level descriptors only. Subjects were urged throughout the experiments to put down their conjectures about the task.

All verbalizations from the protocols were automatically recorded by the software used. Five types of entries constituted a protocol; questions, conjectures and personal notes of subject; and answers and notes by the experimenter.

II. RESULTS AND DISCUSSION

The protocols have provided direct and indirect accounts of three kinds of knowledge sought at the onset. (Akin and Reddy, 1977) Aside from these, three things have been accomplished by the experimental method used. One is the ability of bredth-first exploration of the problem space. Unlike other studies -- eye fixation studies, specific task environments with simple visual stimuli -- a broad base of issues of visual processing are tackled, simultaneously. This enables the acquisition of a general view of a large problem space and the cross-cultivation of the knowledge about all major issues being explored.

Secondly, the very fast process of visual perception is slowed significantly enabling the subjects to generate richer data. The paradigm developed here is intended to aid model building in artificial intelligence more so than exploring the issues of cognitive psychology. Therefore, the fact that it places the natural process of visual understanding into a form of problem solving does not present a problem. Finally, the slowed down process of unraveling the scene is channeled through the experimenter, enabling a rich amount of data to be recorded.

In addition to the general scene understanding task reported above, various other tasks have been tried using the same experimental paradigm: finding a landmark (target) in a scene; navigating the experimenter on a path in a scene; and detection of change between two scenes with similar contents.

Finding a landmark (target) in a scene: The subjects are briefed on a map (Figure 3) of the area contained in the scene. They are told what the scene contains and are required to locate and identify a specific target in it. (Table 1) The same kinds of Feature Extraction Operators and Rewriting Rules have been observed in this task as in the original picture-puzzle task. However, the Flow of Control reflects unique patterns Special knowledge sources for translating two different representations of the same scene (from photograph to the map and vice versa) into one another.

Navigating the experimenter on a path: The subjects are briefed about what the scene contains and are required to find a path for navigation around an obstacle. The scene used was a suburban house scene and the obstacle was the house itself.(Figure 4 of the first paper in this volume, entitled "Knowledge Acquisition") Here special knowledge sources for translating the functional requirements of navigation into spatial terms are used in the protocols. (Table 2)

Detection of change in two scenes with similar content: This experiment aims to simplify the original task eliminating detailed examination of the scene all together. Instead of requiring subjects to determine the nature and the contents of a scene the task requires subjects to match two photographs with slightly different contents. For example, the subject is told that there are two photographs: one representing a central business district of a large city (Figure 2 of the first paper in this volume, entitled "Knowledge Acquisition") and the other representing an urban industrial sector of the same city (Figure 4). The task was to identify each photograph based on this distinction. (Table 3) This task enabled the exploration of only a subset of the original task, i.e., discovering the nature of the scene, independent of a detailed exploration of the scene's contents.

The experimental paradigm explored here provide new means of exploring the knowledge acquisition process in image understanding tasks . We have cited some variations of the paradigm above. These examples however are not exhaustive of all of its possible uses.

III. ACKNOWLEDGEMENT

We thank Keith Price for the valuable help he provided by programming TALKK that enabled the automatic recording of the protocols in the picture-puzzle experiment.

REFERENCES

1. Akin, O. and Reddy, R. Knowledge acquisition for image understanding research, *Journal of Computer Graphics and Image Processing*, 1977 (in print).

2. Baylor, G. W. A Treatise on the Mind's Eye: an Empirical Investigation of Visual Mental Imagery. Ph.D. Thesis, Carnegie-Mellon University, 1971.

3. Buswell, G. T. *How People Look at Pictures*, University Press, Chicago, 1935.

4. Farley, A. M. VIPS: A Visual Imagery and Perception System: the Results of a Protocol analysis, 2 vols. Ph.D. Thesis, Carnegie-Mellon University, 1974.

5. Loftus, G. R. A Framework for a theory of picture recognition, presented at the National Academy of Sciences Specialist's Meeting on Eye Movements and Psychological Processes, Princeton, N. J., April, 1974.

6. Mackworth, N. H. and Morandie, A. J. The gaze selects informative details within pictures, *Perception and Psychophysics*. 1967, 2, 547-551.

7. Moran, T. P. The symbolic nature of visual imagery, *Proceedings of the Third International Joint Conference on Artificial Intelligence*, Stanford, CALIFORNIA, 1973.

8. Ohlander, R. B. Analysis of Natural Scenes, Ph.D. Thesis, Carnegie-Mellon University, April, 1975.

9. Potter, M. C. and Levy, E. I. Recognition Memory for a rapid sequence of pictures, *Journal of Experimental Psychology*, 1969, 81, 10-15.

TABLE 1 Sample Protocol from Task 2.

SUB: READY
EXP: find the intersection of liberty and commonwealth pl. in the picture frame

(1) SUB: PLEASE TELL ME THE LOCATION OF THE VISIBLE PORTION OF THE RIVER (ELONGATED BLUE STRIP).
NOTE: I passed on a tracing, no.1 of the outline of river
EX' go

(2) SUB: IS THERE ANOTHER ELONGATED BLUE STRIP IN THE LEFT HALF OF THE PICTURE.
EXP: yes

SUB: PLEASE TELL ME THE LOCATION OF THAT ONE TOO
NOTE: tracing no 2 of sky
EXP: go
CONJECTURE: NEW BLUE REGION IS SKY.

(3) SUB: IS THERE A SMALL BLUE REGION, NOT NECESSARILY ELONGATED, BELOW OR TO THE RIGHT OF REGION 1.
EXP: yes there are 3-4 narrow blue bands below 1

SUB: PLEASE SHOW ME THE LOCATIONS OF THE BLUE BANDS.
EXP: go

SUB: IS THERE A
THIN NON BLUE STRIP GOING ACROSS REGION 1, ROUGHLY
PERPENDICULAR TO THE LONG AXIS OF REGION 1.
NOTE: subject asked about the exact location of camera. sketch no. 3 is passed on
EXP: no
CONJECTURE: REGION 1 IS NON INTERRUPTED HOMOGENEOUS COLOR.
MAYBE THE FT. DUQUESNE BRIDGE IS IMMEDIATELY TO RIGHT OF REG.1

SUB: IS THERE ANOTHER BLUE REGION TO THE RIGHT OF REGION 1,
BUT SEPARATED FROM REGION 1 BY SOMETHING ELSE.
NOTE: explained to subject that all regions given are noninterrupted homogeneous colors
EXP: no

(4) SUB: IS THERE A QUADRILATERAL SHAPED REGION IMMEDIATELY TO THE RIGHT OF REGION 1.
EXP: yes

SUB: PLEASE SHOW ITS COORDINATES.
NOTE: sketch no.4
EXP: go

(5) SUB: ARE THERE SOME THIN LIGHT-COLORED STRIPS BELOW REGIONS 1 OR 4.
EXP: yes they are quite thin

SUB: PLEASE DRAW THREE OF THEM FOR ME.
NOTE: sketch no 5 miscellaneous roads, bridges and sidewalks
EXP: go

(6) SUB: IS THERE ANOTHER QUADRILATERAL SHAPE IMMEDIATELY TO RIGHT OF REGION 4. IF SO PLEASE DRAW IT FOR ME.
EXP: no
CONJECTURE: MAYBE ANOTHER BUILDING OCCLUDING PART OF THE OTHER SIDE OF THE REG. 4 BUILDING

SUB: PLEASE DRAW THE REGION WHICH TOUCHES REGION 4 ON THE RIGHT.
EXP: go
CONJECTURE: FROM THE RELATIVE WIDTH OF TWO PARTS OF
REGION 4 (I.E. RIGHT PART IS WIDER)
REGION 4 IS PROBABLY BUILDING 96 ON THE MAP OR BUILDING 65 OR
51. NO CANT BE 51 BECAUSE 51
THE REGION 5 THIN STRIP RIGHT
BELOW THE LEFT PART OF REG 4 IS COMMONWEALTH AVENUE.
I NOW HAVE A GOOD IDEA OF THE SCALE OF THE
BUILDINGS. BIGGER (IN THE PICTURE) THAN I WAS EXPECTING BEFORE.

SUB: PLEASE DRAW THE REGION THAT FITS INTO THE CORNER OF THE REGION YOU JUST DREW.
NOTE: drew no.6 and drew no.7, facade of Hilton.
EXP: go
CONJECTURE: BUILDING 7 SMALLER THAN BUILDING 4-6 SO PROBABLY NOT BUILD
65 AND 51 RESPECTIVELY IN THE MAP.

SUB: ARE THERE SOME THIN NON BLUE STRIPS APPROXIMATELY. COLINEAR WITH THAT REGION 5 STRIP RIGHT BELOW REG. 4 IF SO PLEASE DRAW THEM.
EXP: no

(7) SUB: ARE THERE SOME GREEN REGIONS BELOW REGIONS 4, 6 OR 7. IF SO PLEASE DRAW THREE
NOTE: drew some trees and bushes that was drew no.8.
EXP: go

(8) CONJECTURE: THE ACCESS ROADS BETWEEN FT PITT BRIDGE AND
FT DUQUESNE BRIDGE ARE NOT EVEN IN THE PICTURE. OFF TO THE LEFT
AND BELOW PICTURE. I FEEL RATHER STRONGLY THAT BUILDING 4-6
REGIONS 4 AND 7 SHOULD BE ROUGHLY COLINEAR
WHICH THEY ARE NOT IN MY IMAGE. SO I'M STILL CONFUSED.
A BIT. THE LONG THIN BREAKS BETWEEN THE GREEN
REGIONS ARE PROBABLY THE ROADS I'M LOOKING FOR I.E.,
LIBERTY AND THE TWO SECTIONS OF COMMONWEALTH CUT BY LIBERTY.

TABLE 2 Sample Protocol from Task 3.

SUB: WHERE IS THE HOUSE IN THE PICTURE?
EXP: I am incapable of determining identity of semantic things like a house, trees etc.

SUB: OK. WHERE ARE THE GREEN AREAS OF THE PICTURE?
NOTE: drawing 1 the green areas.
EXP: go

SUB: ARE THERE ANY BLACK OR GRAY AREAS CONTAINED WITHIN THIS GREEN AREA?
EXP: there are some gradual darkening of the color at some areas which sometimes is as dark as black
CONJECTURE: GREEN FOR GRASS, WHICH MIGHT BE USED FOR THE PATH AROUND THE HOUSE
GRAY OR BLACK FOR WALKS, ETC.

SUB: WHAT IS THE PREDOMINANT COLOR OF THE REGIONS DIRECTLY ABOVE THE GREEN REGION?
EXP: brown

SUB: COULD YOU OUTLINE THE BROWN REGIONS?
NOTE: drawing 2
CONJECTURE: BROWN FOR HOUSE? STRAIGHT LINES ON PART OF THE BROWN AREA APPEAR TO BE MAN MADE OBJECTS I.E. A HOUSE.

SUB: WHAT IS THE COLOR OF THE REGION DIRECTLY ABOVE THE EXTREME LEFT SIDE OF THE GREEN AREA?
EXP: blue

CONJECTURE: BLUE = SKY (USUALLY) AND MEANS THAT THIS MIGHT BE A LIKELY PATH
SINCE THERE ARE NO OBSTRUCTIONS.

SUB: DOES THIS BLUE REGION EXTEND TO THE TOP OF THE PICTURE?
EXP: yes
CONJECTURE: MUST BE THE SKY

SUB: DOES THIS REGION (BLUE ONE) ALSO CONTACT THE GREEN ONE ON THE RIGHT SIDE OF THE PICTURE?
EXP: yes
CONJECTURE: THIS MIGHT GIVE THE EXTREMES OF THE OBJECT
BETWEEN THE GREEN (PROBABLY GRASS ETC) REGION
AND THE SKY.

SUB: COULD YOU OUTLINE IT?
EXP: the blue region?

SUB: YES.
NOTE: drawing 3, the blue region
EXP: go
CONJECTURE: LOOKS LIKE A HOUSE.
SINCE THE BORDER BETWEEN THE BLUE AND THE
GREEN IS STRAIGHTER ON THE LEFT
THAN THE RIGHT IT LOOKS LIKE THIS MIGHT BE A
BETTER CHOICE FOR A ROUTE AROUND
THE HOUSE. THE HUMPS IN THE BORDER ON THE
RIGHT ARE PROBABLY SHRUBS.

SUB: CAN YOU OUTLINE THE MORE OR LESS VERTICALLY ORIENTED GREEN REGIONS?
EXP: do you mean green regions that are elongated and vertical in orientation?
If so then the green regionive given you is one continuous green
And it is not elongated in the vertical direction.

SUB: NO. THE ORIENTATION OF THE SURFACE I.E. APPROXIMATELY. CONSTANT DISTANCE FROM THE OBSERVER.
NOTE: drawing 4, vertical (3-d) greens
EXP: go
CONJECTURE: THESE SHOULD BE REGIONS TO AVOID SINCE THEY ARE
USUALLY TREES OR SHRUBS
OR POSSIBLY HILL SIDES.

SUB: IS THE GREEN REGION ON EXTREME LEFT RELATIVE FAR OR NEAR COMPARED WITH THE HOUSE (THE VERTICAL ONE)?
EXP: far
CONJECTURE: FAR FROM THE HOUSE, SO IT WILL NOT INTERFERE WITH
THE PATH.
TO GET AROUND THE HOUSE YOU SHOULD STAY ON
THE GREEN REGIONS,
BUT ONLY THOSE THAT ARE
NOT VERTICAL. AND YOU SHOULD AVOID THE BROWN
REGIONS (PROBABLY THE HOUSE)
THE BEST PATH IS TO THE LEFT SINCE THE ONE TO THE
RIGHT IS BLOCKED BY SHRUBS (PROBABLY)

SUB: SEE PICTURE 1 FOR PATH.

TABLE 3 Sample Protocol from Task 4.

EXP: Find which one of the pictures I have, the first one or the second one, is of a downtown area and which is of an industrial, urban area?

SUB: Are there any large rectangular areas in the scenes?
EXP: Yes there are.

SUB: I could find the sky finding operator or grass finding operator but it's hard to find a warehouse or skyscraper finding operator. OK. so let's look at that big rectangular region in the first one. What's it typed with.
EXP: This is about...you want dimensions?

SUB: No ratio
EXP: OK. It's width is one fifth the width of the picture. Its height

SUB: I meant ratios to each other
EXP: Ratios to each other. Seven to ten

SUB: And how about the large one in the other one?
EXP: One is about

SUB: You know I'm kind of matching corresponding parts.
EXP: Let's in like seven to ten.

SUB: Would you classify the texture in picture 1 in that region as high, moderate, or low?
EXP: How would you measure...is the degree of contrast between differing areas?

SUB: Within itself
EXP: Within itself. Do you mean contrasts? I don't have any internal measurement texture so I was trying to get yours.

SUB: This would oe a untextured from a distance.
EXP: Yes.

SUB: OK. Let's say busy. Are there a lot of little regions?
EXP: Yes. It is busy.

SUB: If you were to look closely at that first region would see a lot of little regions?
EXP: Yes.

SUB: And would you say the same thing on the second texture? What's the basic color of that in picture 1 and 2?
EXP: Picture 1 it is basically gray. The other one is brown.

SUB: I guess that in the first one that it is probably a cement building and the second one would be a brick building and the cement building texture could be caused by a lot of windows or something like that and in the second one, brown would be caused by brick. If we were to look closely at...Oh! Let's look at the quality of that business. Is it regular in number 2 and ... is it regular in each one?
EXP: Yes

SUB: That supports the brick window theory. You can only give me relative range, is that it?
EXP: Yes. No absolute range

SUB: I wish I had a physical size. Warehouses can be cement or brick office buildings can be cement or brick, OK let's look at the regions touching these regions. OK. How many regions adjoin that region in the first picture?
EXP: Well, 1,2,3,4, I'd say about 10 to 15 in the first picture. Five or six in the second.

SUB: OK. In the second let's get the biggest. How many did you say in the first?
EXP: Ten to fifteen

SUB: Since there are fewer in the second let's start with the big one. Look at the big one in relation to the regions mumble.
EXP: It is to the right.

SUB: And what is its size?
EXP: What size? Its proportion?

SUB: The proportion to anything.
EXP: It is roughly twice the initial mumble.

SUB: And what's its color?
EXP: It is gray. Black.

SUB: Does it border on the whole side? We have one side there the right side of the region we are talking about. That region that is touching it. Does it touch the whole side of that?
EXP: It only touches it partially.

SUB: Does it touch the top side of it, some side or irregular, top right?
EXP: The right hand side. This is the right side.

SUB: Then we have the right side. Does it touch the whole right side? Or just part of the right side?
EXP: Oh. The bottom, bottom.

SUB: Does it touch any of the bottom of the picture?
EXP: No.

SUB: Backing up a little. Picture 1 that region in question is one sixteenth and the region in question in picture 2 is one hundredth.
EXP: Yes.

SUB: What was the size of this region?
EXP: Twice.

SUB: Twice this guy.
And what is his shape? Is it...
EXP: It is a five sided shape. Make it seven.

SUB: On a regular on the seven sided object?
EXP: Roughly. It's not actually.

SUB: So these seven sides, its regular and then all sides are about the same?
EXP: No

SUB: It's irregular?
EXP: It's irregular

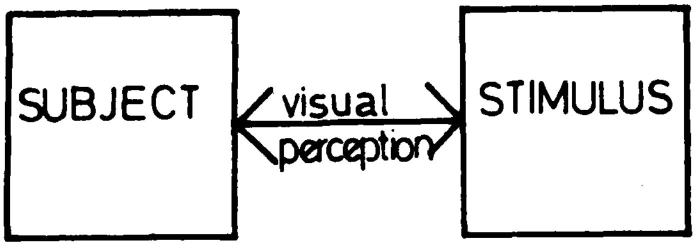


FIGURE 1 Information Flow in Visual Image Understanding.

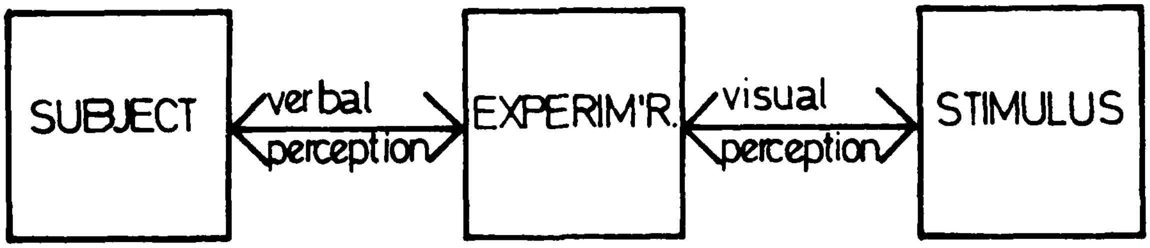


FIGURE 2. Information Flow in the Picture-Puzzle Task.

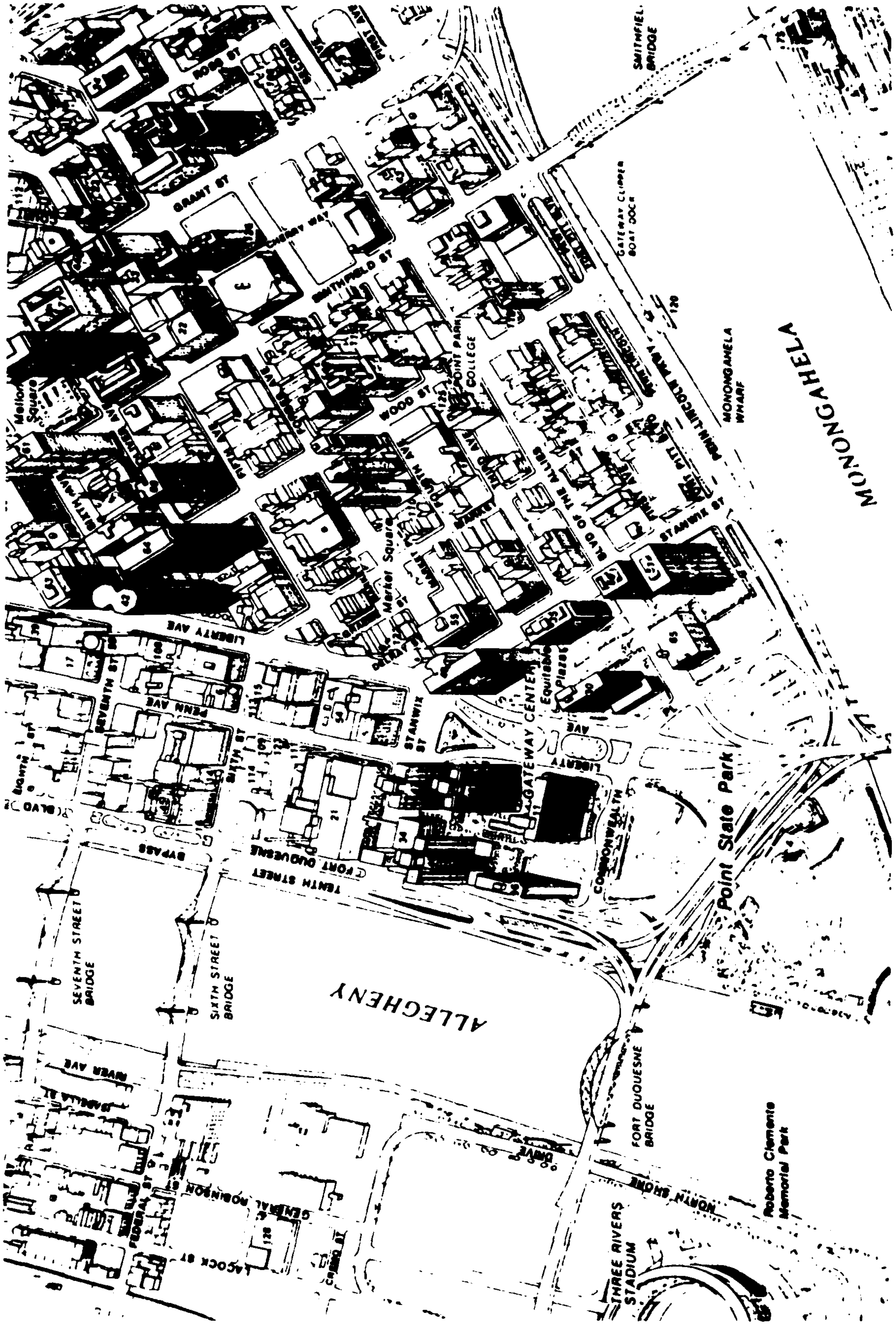


Figure 3. Map of Downtown Scene

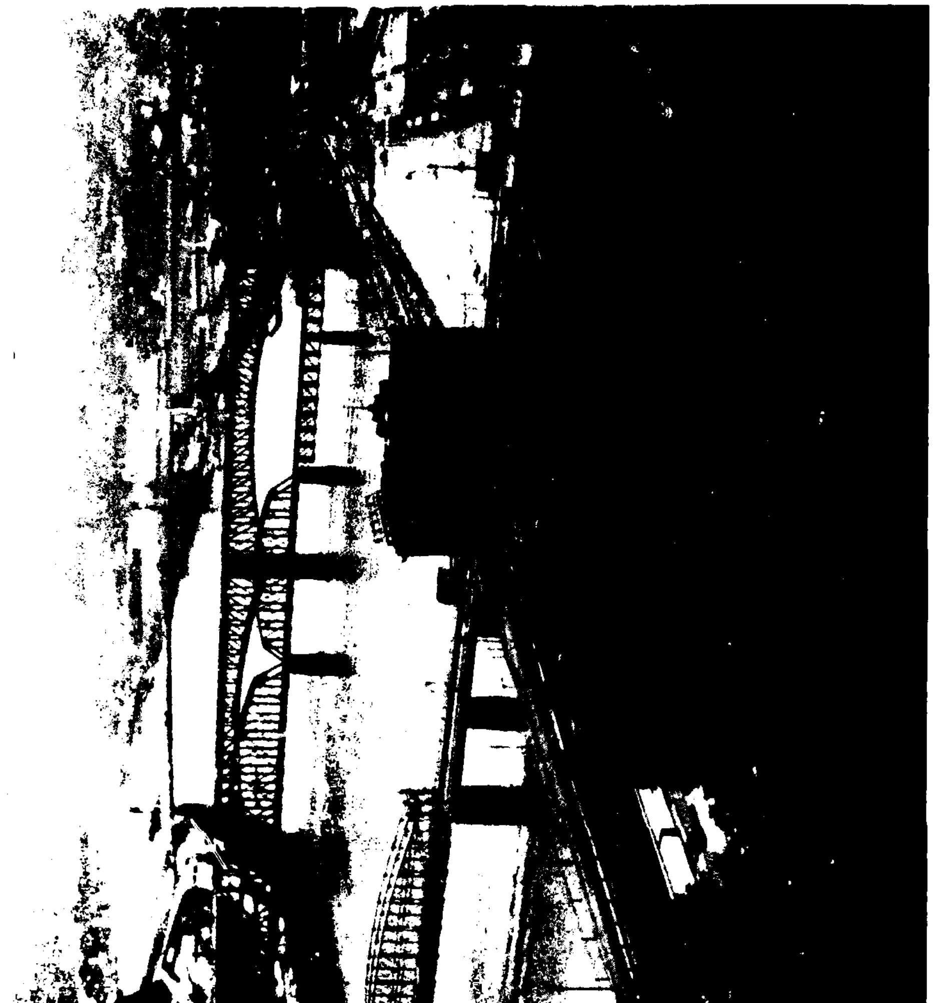
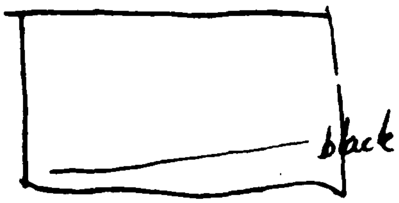


Figure 4. Industrial Area Scene

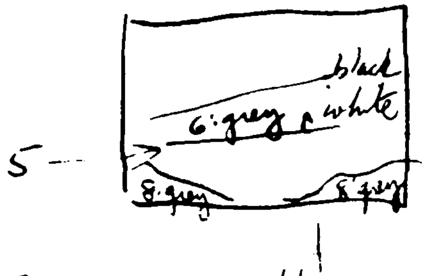
1) mostly shades of green ; varies slightly : little brown patches

2) three definite lines : black , white ; grey

3)

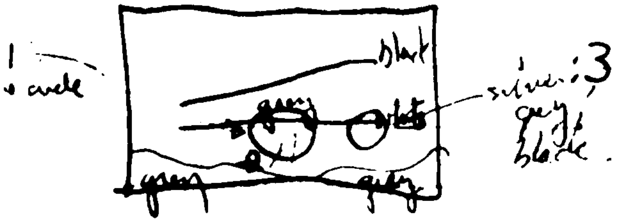


4



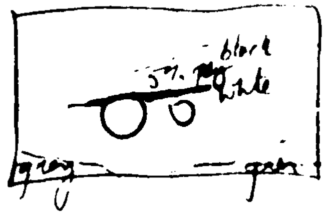
7: grey line not there

rash!



2: silver grey black

1 :



5

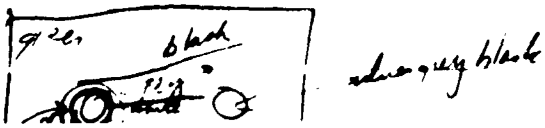


Figure 5. Notes of Subject 1

An Interactive Protocol Analysis System for Knowledge Acquisition

Omer Akin* and Marty Schultz
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pa. 15213

An experimental paradigm for exploring the use of knowledge in understanding images has been developed by Ohlander, Reddy and Akin. (1976) The experimental tool used yields a verbal protocol of subject behavior during the "picture-puzzle" task. This paper describes an interactive computer program that aids the trancription of the protocols obtained.

In the picture-puzzle task subjects are required to determine the contents of a color photograph (Figure 1 in the first paper of this volume, entitled "Knowledge Acquisition") without ever seeing it but by asking questions about it to the experimenter. The experimenter answers all questions about the photograph that do not involve high-level concepts and objects. The only information given about the photograph is low-level information like shapes, colors, locations, textures of different regions in the photograph. The protocol consists of all conversation that takes place between the subject and the experimenter. (Table 1 in the first paper of this volume, entitled "Knowledge Acquisition")

Protocol analysis has been used by Newell and Simon (1972) and later by others (Eastman, 1970; Baylor, 1971; Farley, 1974) to analyze similar verbal data. Even though Waterman and Newell (1973) have developed an automated protocol analyzer, their system is not suitable for our needs. In this paper we present a framework and an interactive computer aid for the analysis of protocols obtained from the "picture-puzzle" task.

The objective of the protocol analysis in this study is to identify the ***knowledge sources*** used in the picture-puzzle task. The categories of knowledge sought are three-fold; Feature Extraction Operators, Rewring Rules, and Elements of Control Flow. (Akin and Reddy, 1977) The categories identified with ease in the analysis are the Feature Extraction Operators and the Rewriting Rules. The protocol analysis also provides some insight into the kinds of Control Elements used in the task.

Feature Extraction Operators: Subjects doing the picture puzzle task use a variety of descriptive terms to identify those features of objects necessary for recognition. These terms cover the categories; scene description, size, shape, color, texture, location, quantity, representational, patterns and miscellaneous others.

* Also in the Department of Architecture.

Rewriting Rules: Some production-like rules have been used by the subjects, mostly implicitly, in order to translate the low-level scene descriptors into high-level concepts or objects. Some examples are: "green indicates grass," "gray, linear, and horizontal surfaces indicate roads."

Elements of Control Flow: Subjects generally used a ***hypothesize and test*** strategy. Other specific strategies were also employed to generate the next hypothesis, apply the next test, and determine the next issue to be explored in special task contexts.

I. A FRAMEWORK FOR THE PICTURE-PUZZLE TASK

A primary objective in protocol analysis is to identify the ***problem states*** and the ***operators*** that are used to move the current *task state* closer to a ***solution state*** incrementally. The protocol analysis system used here tries to do the same. A set of Task Operators have been defined *a priori*. Some of these Operators applied in the picture-puzzle task are identified automatically using prior knowledge about these operators and others are identified manually by the experimenter in an interactive mode.

Three macro Task Operators have been consistently observed in all protocols. These are: 1) *Search*; select an issue or aspect of the scene to explore, 2) ***Hypothesize***; generate an hypothesis about the identity of the issue(s) being explored, and 3) ***Test***; apply appropriate tests to clarify the hypotheses generated. All three kinds of knowledge defined above are used in the Search, Hypothesize and Test Operators.

For example one of the subjects uses the knowledge that "scenes can be classified into two in general; outdoors and indoors" to ***select*** the first issue to deal with. Then he *generates* a hypothesis (i.e., outdoors) based on the same knowledge. Later he tests the converse hypothesis as well (i.e., indoors). In ***testing*** the "outdoor" hypothesis he uses the Rewriting Rules that "outdoor scenes contain a part of the sky" and "sky is blue." After both tests fail (i.e., neither outdoor or indoor scene) the subject goes back to the above Rewriting Rules and modifies them to read: "outdoor scenes contain a part of the sky, unless the sky is completely occluded by other objects," and "overcast skies are gray." This leads to the correct resolution of the issue, i.e., the scene is an outdoor scene with occluded sky.

II. TRANSCRIPTION OF PROTOCOLS AND ANALYSIS

Identification of the Feature Extraction Operators requires manual search of the text for terms describing some visual aspects of the scene or some of its parts. Identification of the Rewriting Rules requires the determination of what new information is acquired by the subject in each state and what Rewriting Rules are being applied to translate all the accumulated information into an assertion about the scene. Finally, in order to identify the Elements of Control Flow a transcription of the protocol into a form in which patterns of search are clearly seen is needed. The most proper format for achieving this is the Problem Behavior Graph used by Newell and Simon. (1972)

Each protocol consists of questions asked by subjects, answers given by the experimenter, and comments made by subjects (both conjectures and notes made by the subjects about their own behavior). The task of the protocol transcription system presented here is to take this information and aid the analysis in coming up with the three kinds of knowledge used in each Task Operator: Search, Hypothesize and Test. A sample of a transcribed protocol is provided in

Table 5 in the first paper of this volume, entitled "Knowledge Acquisition." This sample corresponds to the first seven question-answer sequences of the sample protocol.

The protocol analysis system (PROTDO) was developed* to simplify the manual task of the human transcriber. PROTDO performs four major operations. First it gets the file of the protocol to be transcribed. Next, it displays each question-answer sequence along with the previous and the next question-answer sequences in the protocols. Then, it allows the transcriber to enter all Task Operators and related knowledge sources for each question-answer sequence being transcribed, individually into the transcribed file. While doing so PROTDO stores each question-answer sequence along with the knowledge entered for each Task Operator. Some knowledge sources, such as Feature Extraction Operations, are built into the "memory" of PROTDO. This enables PROTDO to automatically identify some knowledge sources. Finally PROTDO stores all this information in a new file before quiting on the protocol being worked on.

III. HOW TO USE PROTDO

The first question a potential user should ask himself is "do I really need to use PROTDO"? Because PROTDO is a program especially tuned to the transcription of protocols taken with the picture-puzzle task and with the objective of discovering the knowledge sources outlined earlier. Transcriptions with different intent and/or other task protocols are very likely to be unsuitable for PROTDO.

When PROTDO is run, first it will ask the user if he needs help with the program. If yes "Y" is replied a brief summary of program usage is printed. Next the user is asked the file name to be processed, followed by the file name to store the transcribed protocol in. Next, the number of the protocol to be transcribed is requested (multiple protocols can be stored in a single file, each delimited by a page mark).

PROTDO then asks for a file name to store the set of Rewriting Rules (RR) under, and a file name for the collection of Elements of Control Flow (ECF). To avoid creation of either file, the user presses the return key without typing a name to the respective prompt.

Now PROTDO can start to process the protocol selected. First, PROTDO displays the previous question-answer sequence just processed along with the present sequence on the CRT. In this fashion PROTDO displays all question-answer sequences in pairs until the end of the protocol.

* PROTDO has been programmed by Marty Schultz.

The processing of each question-answer sequence displayed consists of entering all knowledge sources used for each of the Task Operators in that sequence. That is for each of the three Task Operators *search*, *hypothesize* and *test* all three types of knowledge sources are sought, i.e., Feature Extraction Operators, Rewriting Rules and Elements of Control Flow. The previous question-answer sequences is displayed with each current question-answer sequence to enable the user to see the context of the current sequence.

After PROTDO has displayed the appropriate sequence of questions, the user can enter one of three commands. A slash "/" instructs the program to terminate interactive analysis, and finish writing the files using only that which has already been processed. A star "*" causes PROTDO to ignore this sequence and go on to the next one. Any other character begins interactive analysis of the present sequence.

The first thing PROTDO does after encountering a character other than a "/" or a "*" is to display the keyword "SEARCH" as the first category of Task Operators. At this point the user has to decide what issue is being dealt with in the current question. Then the user has to type in the issue being dealt with and return control to PROTDO. This will cause PROTDO to save that entry as the description of the *search* Operator of the current question.

The other Task Operator categories, *hypothesize* and *test*, are processed similarly. That is a keyword is prompted and the user enters a *hypothesis or test* description. PROTDO automatically proposes the text of the question asked by the subject as the description of the *test* Operator. The user can accept this description by typing "Y" for yes, anything else for no. If it is rejected PROTDO will expect the user to type in a *test* category description just as in the previous two categories of Task Operators.

Right after successfully entering any of the three Task Operator descriptions, PROTDO enables the user to enter descriptions of the three classes of knowledge sources; Feature Extraction Operators (FEO), Rewriting Rules (RR) and Elements of Control Flow (ECF). PROTDO first displays the appropriate keyword for each knowledge source category, i.e., FEO, RR, ECF. For each keyword PROTDO expects the user to either accept the description it provides automatically or to enter a new description.

PROTDO has a memory consisting of all FEO's it has ever encountered. Every time a new FEO is entered in a transcription file PROTDO saves it in its memory for future transcriptions. Hence whenever the FEO category comes up during a transcription session PROTDO finds words in the Task Operator description that match FEOs in its memory and displays these on the CRT along with the keyword "FEO." When PROTDO chooses the Operators, the user can edit these choices.

As each FEO is printed, the user can accept it by typing a comma or a period. The comma will cause the FEO in the final transcription to be separated by a comma. A period requires the use of a blank as the separator. This latter choice is used in multiple word FEOs, such as "with respect to." Any other character typed will reject that FEO for this sequence. After all operators have been generated, PROTDO will ask if any others (not in its dictionary) are to be included. If the user wishes to enter more, he types them here, each delimited by a blank or a comma. Otherwise the user hits the return key. This will commence the entry of the FEO description. The FEOs added here are subsequently combined in PROTDO's dictionary upon program exit.

By the time all three Task Operators are processed the information entered on the CRT will have been stored in the transcription file along with the text of the current question-answer sequence. After the completion of the last question-answer sequence the transcription file will be closed. As was mentioned before, a slash can be used to terminate transcription before starting the processing of the current question-answer sequence. This will cause PROTDO to save the total transcription completed up to the current question-answer sequence in the transcription file. The Rewriting Rule and Elements of Control Flow files will also be saved, if they were declared at the onset.

IV. CONCLUSIONS

PROTDO is useful for a special kind of transcription, i.e., looking for knowledge sources, in the picture-puzzle task. Consequently its usefulness in the general sense is limited. However, it provides for us a rich catalogue of the knowledge used in the specific area of research.

Furthermore the output of PROTDO can be easily translated into the Problem Behavior Graph format. This is necessary for observing the general patterns of Control Flow. Each task operation included in the transcription represents a modification in the problem state. Hence these are represented as right arrows linking nodes (problem states) in Figure 1. Every time a question-answer sequence does not alter the problem state, that is the task operation is the same as the previous one, the down arrows are used to indicate no advance in the problem state. The links starting from earlier nodes indicate backtracking which correspond to going back to an issue dealt with earlier in the transcription. The Problem Behavior Graphs obtained from different tasks is expected to yieid a more parsimonious understanding of the Elements of Control Flow.

REFERENCES

1. Akin, O. and Reddy, R., Knowledge Acquisition for Image Understanding Research, *Journal of Computer Graphics and Image Processing* 1977 (in print).

2. Newell, A. and Simon, H., *Human Problem Solving*, Prentice Hall, New York, 1972.

3. Ohlander, R. B.; Reddy, R. and Akin, O., An Experimental System for Knowledge Acquisition in Image Understanding Research, Computer Science Department Report, Carnegie-Mellon University, 1976.

4. Waterman, D. A. and Newell, A., PAS-II: An interactive task-free version of an automatic protocol analysis system, *Third International Joint Conference on Artificial Intelligence*, 1973.

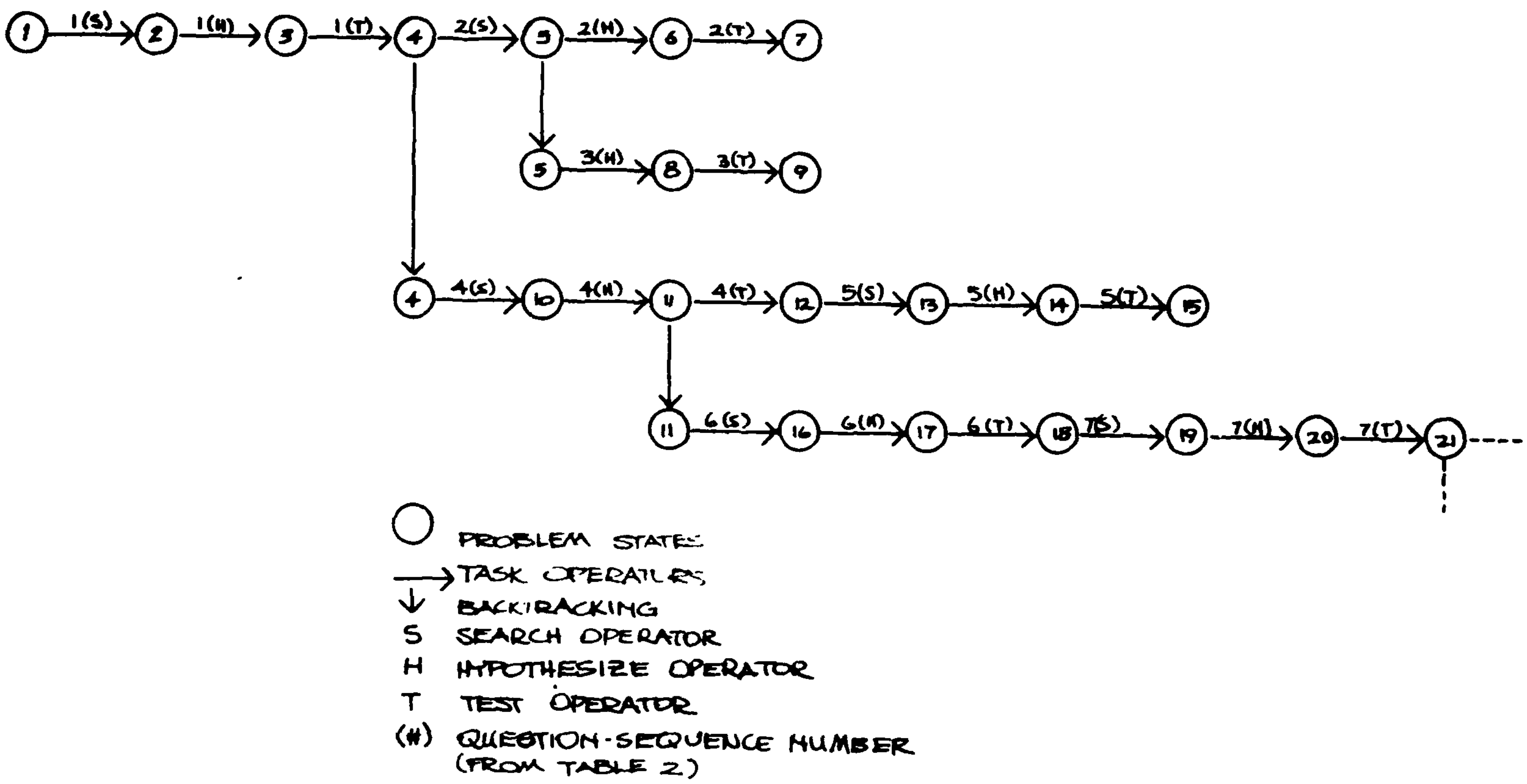


FIGURE 1. PROBLEM BEHAVIOR GRAPH OF TRANSCRIPTION.

Eye Fixations in Image Understanding Research

Omer Akin[*]
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

This study explores an alternative experimental tool for discovering knowledge used in understanding visual scenes. This issue has been examined earlier by Akin and Reddy (1977) using verbal protocols. The possibilities of using visual protocols, i.e., eye fixations, in achieving the same ends will be explored in this paper.

I. EYE FIXATIONS AS MEASUREMENT IN VISUAL INFORMATION PROCESSING

A most frequently asked question in research dealing with visual perception of complex scenes is simply, "How do we perceive pictures?" More specificly this question has taken the form:

> "..how information from a visual scene is encoded?" (Loftus, 1974)
>
> "What does a person do when he looks at a picture?" (Buswell, 1935)
>
> "..[do] key regions exist within pictorial displays.. [and are] some stimuli more important than others within the displays?" (Mackworth and Morandi, 1976)

Alternative experimental means have been used to uncover the visual understanding process. Use of eye fixations in image understanding has been an important research tool. Below we shall review a representative sample of major studies done in the area of visual perception using eye fixations.

One of the earliest and most extensive eye fixation studies was undertaken by Buswell. (1935) His experiments consist of measuring eye fixations of subjects observing various stimuli under different task conditions. The main emphasis of the experiment is the interpretation of eye fixation patterns.

More recently, Loftus (1974) has dealt with the issue of recognition. He has recorded eye fixations and recognition responses of subjects perceiving complex scenes. He has also altered the representation and contents of stimuli to control information transmission.

Mackworth and Morandi (1976) looked at fixations and the judgment of "recognizability" of subjects with two complex stimuli. They have analyzed the data by subdividing the stimuli into 64 equal parts.

[*] Also in the Department of Architecture.

All three studies essentially explore the processes responsible for "understanding" and/or "recognition" of pictures. Yet, all have used different means of pursueing this goal. Here I shall report on the characteristics of these alternative experimental means and what each yields in terms of knowledge in the area.

There are basicly four major experimental means used in these studies. The measurement of eye fixations seems to be the common denominator of all. (Loftus,1974; Buswell, 1935; Mackworth and Morandi, 1976) A second experimental measure used is recognition of a previously seen image. (Loftus, 1974) The third paradigm is the use of subjective ranking of some qualitative aspect of the stimuli by the subjects. (Mackworth and Morandi, 1976) And the fourth experimental means used is the decomposition of stimuli into smaller, or less comprehensive parts. (Mackworth and Morandi, 1976) Below we shall discuss the role of eye fixations in relation to other experimental tools.

Of course the central issue in the use of eye fixation data is just what the fixation corresponds to in terms of cognitive processes. Buswell states the common explanation to the issue in the following terms:

".. the center of fixation of the eyes is the center of attention at a given time... The evidence [provided by fixations] in regard to perceptual patterns is entirely objective, but it furnishes no indication, except by inference, as to what the nature of the subject's inner response to the picture may be." (Buswell, 1935)

Buswell's main concern stems from the large variance in fixation durations --i.e., 3-40 thirtieths of a second. He attempts to explain this variance as a function of stimulus characteristics and stages of the perception process. On the other hand this mere inferential evidence is rather significant. Loftus has suggested that even though the fixation durations in a recognition task vary considerably, the subject's performance is a function of the number of fixations rather than the duration of fixations. This implies that the amount of information acquired during a fixation is more or less constant. Therefore the variance in the duration of the fixation results due to processes other than information gathering that takes place during a fixation -- such as what-part-of-the-picture-to-process-next.

Loftus has also shown that by motivating the subjects to perform better it is possible to reduce average fixation durations without affecting recognition performance. This indicates that some extraneous processes or simply idle time may be responsible for this variance.

The single study which has explored eye fixations most extensively and exclusively is Buswell's "How People Look At Pictures." Location, duration and sequence of fixations have been looked at under various stimulus, subject and task conditions. He has inferred differential picture processing stages as a function of the time dimension and task description, as a function of fixation data.

He found initial fixations to be always shorter than successive ones. This is attributed to the use of central cognitive processing in addition to simple visual processing, as the "understanding" of a picture becomes more detailed and/or more semantic. The evidence provided by Mackworth et.al. and others' (1976; Potter and Levy, 1969; Pollack and Spence, 1968) findings indicate that the first fixations serve a different purpose, namely that of finding out the "gist" of a picture, as opposed to the later ones. Loftus has analyzed also the individual fixations discovering underlying internal perceptual processes. He concludes that in terms of information gathering a fixation performs a standard function independent of its duration beyond the first 100 ms.

Hence there seems to be two major functions of a fixation. The first 100 ms. or so constituting the information gathering and the remainder of the fixation duration deriving from the knowledge about the picture a next target location to fixate upon. (Loftus, 1974) If we assume that the information about the picture is internally represented in a structure isomorphic to a hierarchic structure (i.e., more processing time required for processing more detailed parts of the picture) then it is plausible that the Subjects involved in detailed analysis in the later stages of processing have longer fixation durations.

Buswell found that different task situations, such as simple perception, scanning for target recognition or subjective judgment of picture quality tasks, produced different fixation patterns. This indicates that the information provided by fixation behavior in visual tasks is extremely rich. However there is little theoretical basis for explaining the underlying processes responsible for these differences.

II. EXPERIMENT

A basic problem in all eye fixation studies of picture understanding, is the lack of a general theory of the picture understanding process. With the recognition of this fact, we have done some eye fixation studies using the same images analyzed in the paper entitled "Knowledge Acquisition in Image Understanding" and using the framework developed in the same study. (Akin and Reddy, 1977) Based on the findings of the studies reviewed above we have analyzed the pattern of fixations rather than latencies to infer the search behavior exhibited. The results are inconclusive and have lead to more questions than they have answered. However, we present some of the preliminary findings to expose the state of our research to other interested parties.

The eye fixation experiment consisted of instructing subjects to examine a certain feature, i.e., intersection of two major traffic arteries in downtown Pittsburgh, in a map. (Figure 3 in the second paper in this volume, entitled "An Experimental System") Later subjects were instructed to find that particular land-mark, the intersection, in a photograph of the same area (Figure 2 in the first paper of this volume, entitled "Knowledge Acquisition"). The protocol of the visual search behavior of the subjects were taken by recording their eye fixations. An image of the photograph and fixations were super-imposed on video-tape during the experiments. Two subjects were used in this task. Samples from the protocols of these two subjects are contained in Figures 1 and 2. The consecutive numbers in these figures indicate the sequence of the fixations in each experiment. Note that the numbers also indicate the location of the center of each fixation which was about 1/2" in diameter.

III. ANALYSIS

The patterns obtained in the eye fixation protocols are compared against the issues explored in the protocols of the picture-puzzle experiment. In the picture-puzzle task subjects are instructed to find the same traffic intersection in the photograph of the downtown area after examining the map. But in this case the subjects are not allowed to examine the photograph visually. They are given verbal information about the photograph by the experimenter when they ask for it. This experiment is described in detail in the paper entitled "Knowledge Acquisition in Image Understanding Research." (Ohlander, Reddy and Akin, 1976)

First it should be emphasised that the processes underlying the two experiments are radically different. In the case of the eye fixation experiment the subjects analyze "meaningful" parts of what is visually available in each photograph. While the exact nature of the underlying processes which derive the fixations are still a mystery the general consensus is that fixations represent those parts of the scene which are directly informative for each respective processing stage encountered during the interpretation of the visual image.

On the other hand, the subjects searching for a target in a photograph in the picture-puzzle task seem to construct internal representations of stimuli based on the verbal feedback obtained from the experimenter. Subsequent search of the scene is based on this partial, and at times errorful, representation of the scene. The construction of the internal representation is therefore radically different from the case where the search is based on a complete visual scene, as in the eye fixation experiments.

The initial information explored in the case of the picture-puzzle task about an object, such as a building, usually pertains to a simple descriptive property, i.e., trapezoidal outline(s). While an eye fixation on the same object (the building) readily extracts information (possibly in parallel) about many aspects of that object, i.e., shape, texture, orientation, occlusions, shadows, the environment, etc.

Despite these differences it is possible to observe some parallelism between these two processes. Evidence suggests that successive questions about a single entity in the picture-puzzle experiment extract information about many descriptive aspects, i.e., shape, texture, orientation, etc. (Akin and Reddy, 1977) This is similar to the case of the eye fixation paradigm with the exception that the same information may be obtained in parallel in the latter case.

IV. RESULTS

In the discussion below, we shall compare the patterns of eye fixations against the issues explored by successive sets of questions in the picture-puzzle experiment. For example, the subject in the sample protocol from the picture-puzzle experiment (Table 1 in the second paper of this volume, entitled "An Experimental System") examines first, the river; second, the sky; third, the river; fourth, the buildings; fifth, the roads; sixth, the buildings; seventh, greenery and eighth, the road and the intersection. These actions are respectively numbered in the protocol in the table.

This reflects a characteristic pattern where the subject starts from a familiar object (some thing he can identify readily such as the sky, the river, etc.) in the map and then scans all objects that are expected to lie in the path joining the point of departure to the target object (the intersection). Similar patterns are seen in the fixation data where sets of successive fixations land on the same characteristic objects. For example, consider Figure 1. The first few fixations of Subject 1 (1-6) land around the initial fixation (0) in the center of the scene. Then they successively fall on the river (7-8), the buildings (9-11), the greenery and the roads (12-13), the buildings (14-21), one of the target roads (22-24) and finally the intersection (24-25). The rest of the protocol consists of fixations that appear to repeat this pattern of fixations. This can be attributed to the fact that the subject may want to verify his initial findings by repeating his earlier perceptual actions.

The striking similarity in the sequence of the parts of the scene looked at in each experiment is typical. This does not necessitate that we should get the same results every time. This is obvious if we consider the degrees of freedom there are in finding a path between the target and a randomly selected point of departure of search. However, the results obtained here leads us to believe that the kinds of control exercised in the two experiments examined here are very similar.

This result is intuitively correct. A next fixation is possibly made to add to the current knowledge of the system about the scene, and driven by the goal of finding the target in the photograph. While in the picture-puzzle task each "next" question also serves the same purpose. Hence, with proper aggregation of fixations and questions it should be expected that similar patterns of control can be observed in both experiments.

V. CONCLUSIONS

The eye fixation data indicates one major result. The picture-puzzle paradigm used in the experiment reported earlier is an experimental tool for accurately simulating the actual visual understanding process. This on the one hand supports our experimental assumptions and on the other hand provides a more direct means for exploring the issue of Control Flow in visual understanding.

Ideally, what needs to be done in the eye fixation experiment is to enable the subjects to observe the map and the photograph simultaneously, while the protocol of eye fixations are taken. By recording the patterns of fixations for both stimuli it will be possible to infer more directly the information obtained from the map that directs the flow of eye fixations towards the target in the photograph.

VI. ACKNOWLEDGEMENTS

We are indebted to the Department of Psychology for allowing us to make use of their eye fixation recording equipment. We also thank Patricia Carpenter and Adrian Baer for the valuable protocols they have provided.

REFERENCES

1. Akin, O. and Reddy, R. Knowledge Acquisition in Image Understanding Research, *Journal of Computer Graphics and Image Processing*, 1977 (in print).

2. Buswell, G. T., *How People Look at Pictures*, University Press Chicago, 1935.

3. Loftus, G. R., A framework for a theory of picture recognition, presented at the National Academy of Sciences Specialist's Meeting on Eye Movements and Psychological Processes, Princeton, N.J., Apr. 1974.

4. Mackworth, N. H. and Morandi, A. J., The gaze selects informative details within pictures, *Perception and Psychophysics*, 1976, 2, 547-551.

5. Ohlander, R., Reddy, R., and Akin, O. An Experimental System for Knowledge Acquisition in Image Understanding Research, Computer Science Department Reports, Carnegie-Mellon University, 1976.

6. Pollack, I. and Spence, D., Subjective pictorial Information and Visual Search, *Perception and Psychophysics*, 1968, 3, 41-44.

7. Potter, M. C. and Levy, E. I., Recognition memory for a rapid sequence of pictures, *Journal of Experimental Psychology*, 1969, 81, 10-15.

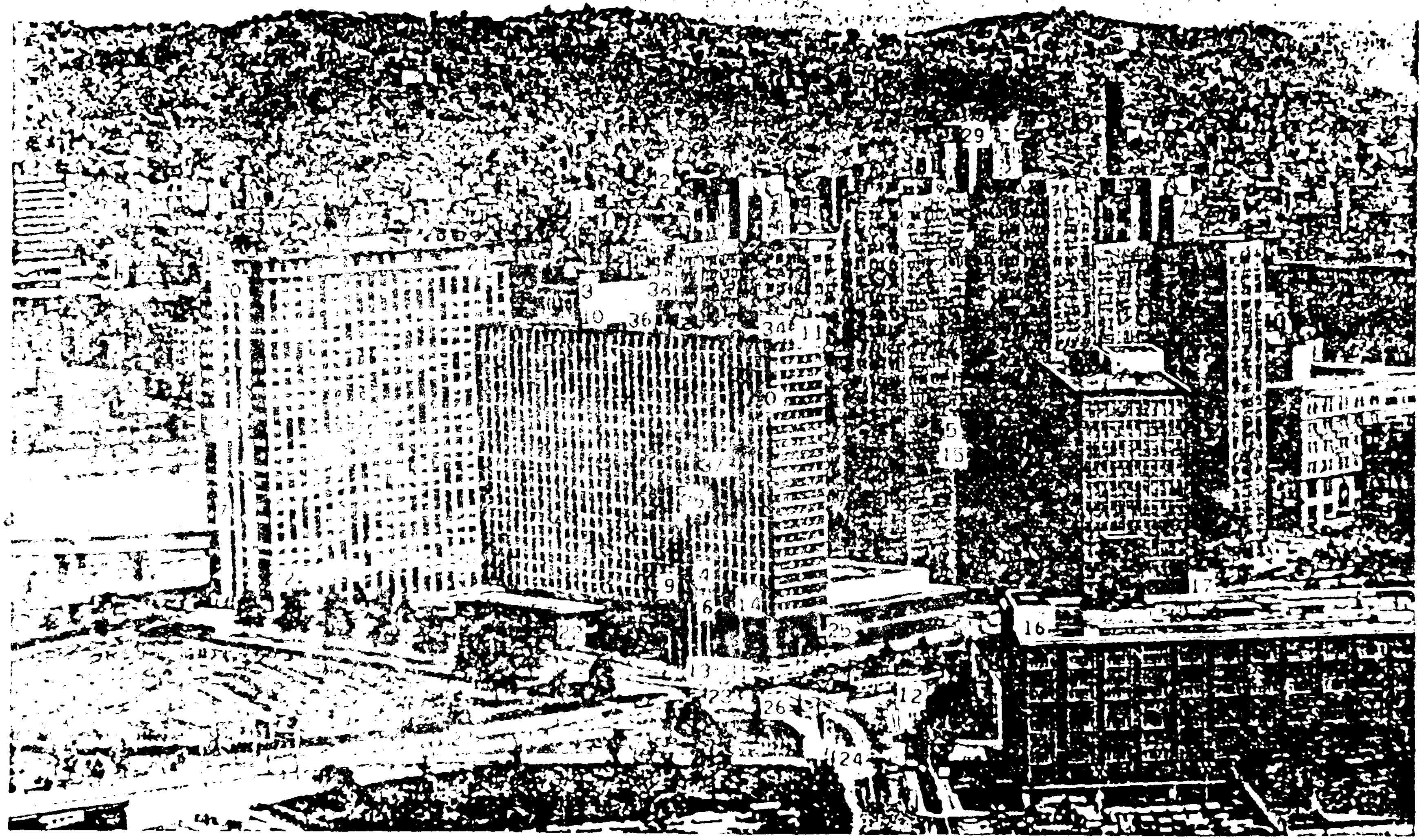


FIGURE 1. EYE FIXATION PROTOCOL OF SUBJECT 1.

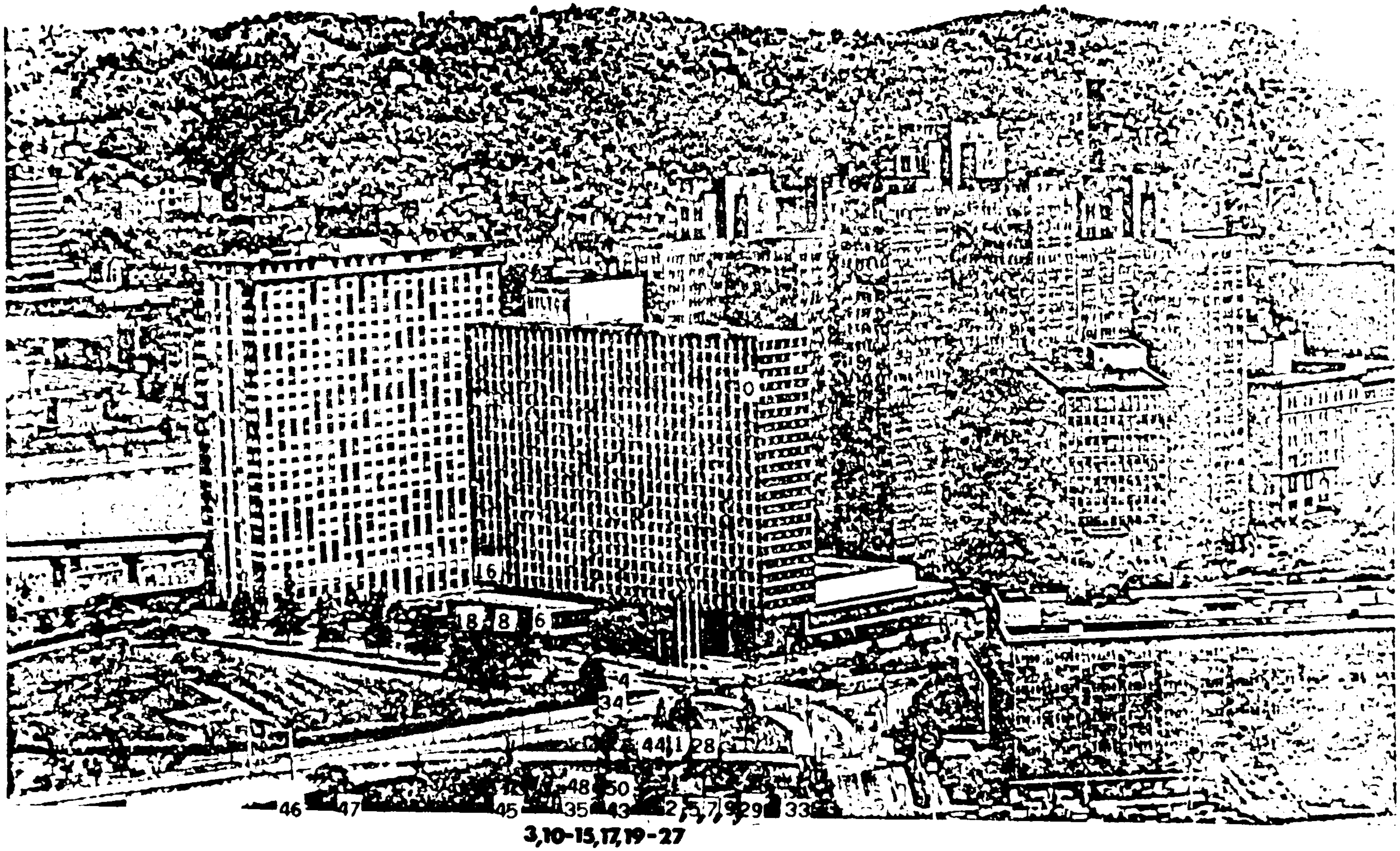


FIGURE 2. EYE FIXATION PROTOCOL OF SUBJECT 2.

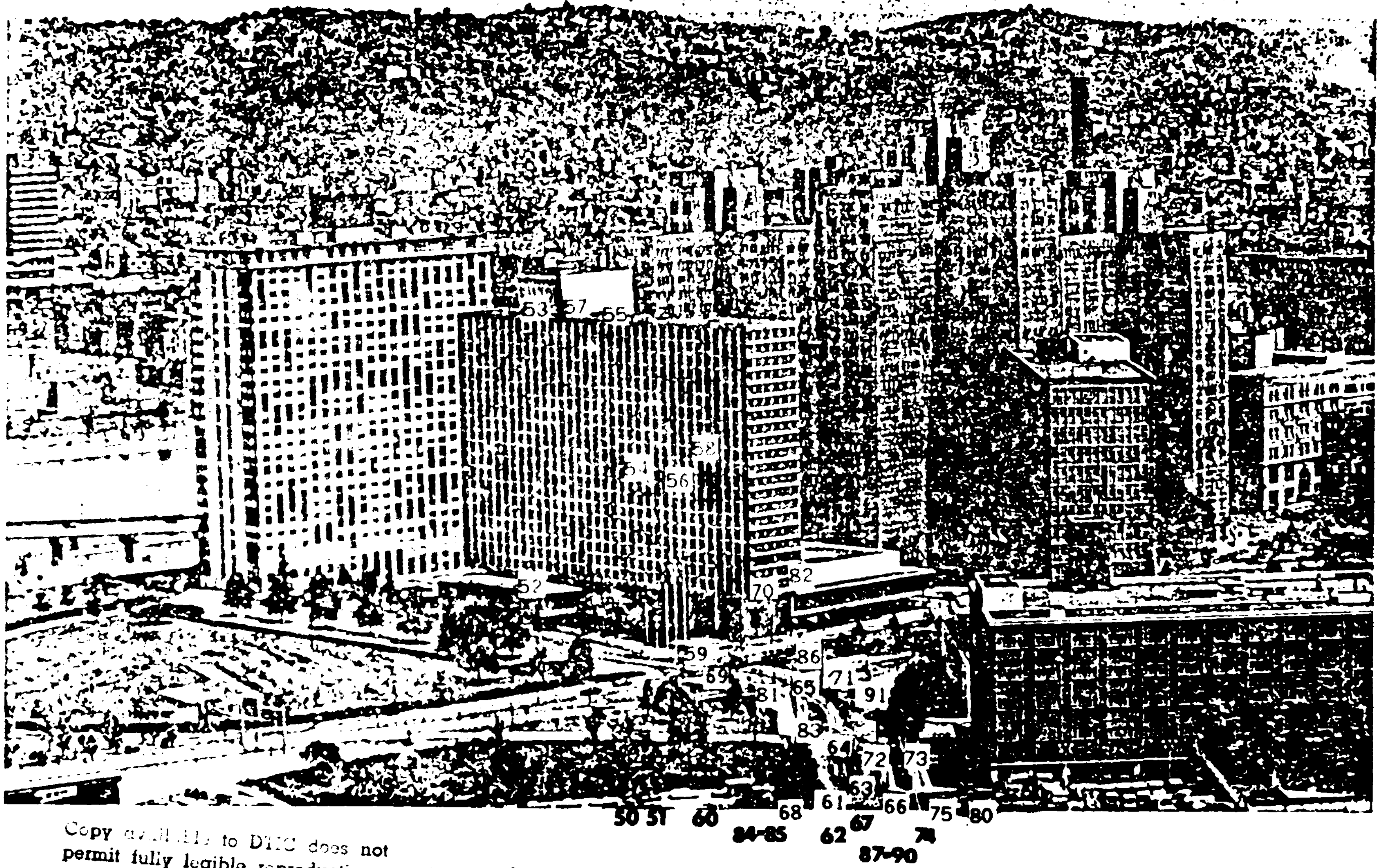


FIGURE 2. EYE FIXATION PROTOCOL OF SUBJECT 2.
Continued.

FILME

9